PB-241 150

PROCEEDINGS NUMBER 1 OF THE OR AND D ADP WORKSHOP HELD IN BETHANY COLLEGE, WEST VIRGINIA ON OCTOBER 2-4, 1974

Denise Swink

Environmental Protection Agency
Washington, D. C.

January 1975

DISTRIBUTED BY:

NTIS
National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. 600/9-75-002	2.
PB 241 150	
4. TITLE AND SUBTITLE OR&D ADP Workshop Proceedings No. 1	5. REPORT DATE January 1975
	6. PERFORMING ORGANIZATION CODE ORD
7. AUTHOR(S) Denise Swink	8. PERFORMING ORGANIZATION REPORT NO. 600/9-75-002
9. PERFORMING ORGANIZATION NAME AND ADDRESS Office of Program Management, Office of Research and Development, U. S. Environmental Protection Agency, Washington, D. C. 20460	10. PROGRAM ELEMENT NO. 1RW103
	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS U. S. Environmental Protection Agency, 401 M Street, S. W., Washington D. C. 20460	13. TYPE OF REPORT AND PERIOD COVERED Symposium Proceeding Oct 2-4, 1974
	14. SPONSORING AGENCY CODE ORD-EPA

15. SUPPLEMENTARY NOTES
The OR&D ADP workshop was held October 2-4, 1974 at Bethany College, Bethany, West Virginia.

16. ABSTRACT
The OR&D ADP workshop was held to meet the need for a better awareness among the EPA Scientific Community, of the available resources, techniques, and applications in automatic data processing within the Agency. Subject areas included; mathematics, scientific, and statistical applications software, minicomputer applications, interactive graphics applications, and laboratory data management.

PRICES SUBJECT TO CHANGE

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Data Processing Computer Systems Programs Statistical Analysis Mathematical Models Information Systems Graphic Methods		07/C

Reproduced by
NATIONAL TECHNICAL INFORMATION SERVICE
Department of Commerce
Springfield, VA. 22151

18. DISTRIBUTION STATEMENT Unlimited	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES 322
	20. SECURITY CLASS *(This page)*	22. PRICE $9.25/2.25

EPA Form 2220-1 (9-73)

600/9-75-002
October 1974

OR&D ADP WORKSHOP PROCEEDINGS NO. 1

Sponsored by

Denise Swink, OR&D ADP Coordinator
Office of Program Management (RD-674)
Office of Research and Development
U. S. Environmental Protection Agency
Washington, D.C. 20460

OFFICE OF PROGRAM MANAGEMENT
OFFICE OF RESEARCH AND DEVELOPMENT
U. S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

FOREWORD

The OR&D ADP Workshop was held October 2-4, 1974, at Bethany College, Bethany, West Virginia. In response to the need for better communication among the scientific community utilizing data processing, the OR&D Workshop was designed to promote the exposing of state-of-the-art data processing techniques and available resources, and the sharing of experience and knowledge related to Agency operating programs involved in scientific data processing applications. The workshop was sponsored by the Office of Research and Development and was open for participation by all organizations within the Agency.

Denise Swink
OR&D ADP Coordinator

TABLE OF CONTENTS

	Page Number
FOREWORD	iii
AGENDA	1
STORET - The Status of the System Charles S. Conger	9
A Method for Improving User Access to STORET Michael J. Friedland	37
Two Data Storage Programs for STORET Kenneth V. Byram	46
The EPA Scientific Applications Software Study Elijah L. Poole	49
BIOMED, SAS, and Other Tools of the Statistician Robert Kinnison, Ph.D.	50
MSP (Model Specification Program)– An Easy-to-Use Linear Statistical Analysis System Larry M. Male	56
Concepts Verification Testing of the ENVIR System John W. Scotton	63
SEAS - The Strategic Environmental Assessment System Edward R. Williams	69
A Computer Controlled Data Acquisition and Retrieval System for Use in Air Monitoring Programs Marvin Hertz, Kirby Kyle, Anne Duke, and George Ward	88
An Aquatic Ecosystem Simulator and VS/8 David M. Cline	96
An On-Line PDP-12 Spirometry Package That Supports Both Onsite and Remote Stations Sam D. Bryan	99
An On-Line Real-Time Multi-User Laboratory Automation System William L. Budde, Edward J. Nime, and Jack Teuschler	104
Computerized Chromatography - Mass Spectrometry William L. Budde and D. Craig Shew	114

	Page Number
CALCOMP Software Theodore R. Harris	118
Harvard Graphics Packages Curtis S. Lackey	132
EPA's Regional Air Pollution Study Robert B. Jurgens	148
Capabilities of Tektronix Software David M. Cline	156
Automated Laboratory Management System in Region V David Rockwell	159
Automatic Data Processing and Regional Laboratory Quality Assurance Billy Fairless, Ph.D.	167
Sample Tracking Data Management System G.C. Allison, M.J. Madsen, and R.N. Snelling	182
SIDES Davis Barrow	194
Selection of Probability Models for Determining Quality Control Data Screening Range Limits Wayne R. Ott, Ph.D.	208
A systematic Approach to Water Quality Trend Analysis Merlin H. Dipert and Jon A. Abraytis	217
Sample File Control for the Automated Laboratory Roman I. Bystroff	219
A Data Reduction System for an Automatic Colorimeter K.V. Byram, F.A. Roberts, and L.A. Wilson	226
A Chemical Information System Stephen R. Heller	234
Mass Spectral Search Systems John M. McGuire	236
Bowne Time Sharing (Word/One) Catherine Tittle	254
Introduction to the UNIVAC 1110 at Research Triangle Park - Capabilities T.L. Rogers	257

	Page Number
RTCC Software and Accessibility Maureen Johnson	265
Technical and Environmental Information System Donald L. Worley	271
The Conversion of CHESS and Other Systems Andrea Kelsey, Gene R. Lowrimore, and Jane Smith	282
Present State of EDP Policy and Development in EPA Theodore R. Harris	289
Status of the Washington Computer Center Procurement Denise Swink	296
APPENDIX A - List of Attendees	A-1
APPENDIX B - Areas of Expertise	B-1
APPENDIX C - Areas of Interest	C-1

WORKSHOP AGENDA

Opening Remarks
Speaker: Denise Swink

Session I
October 2, 1974

Session Title: Mathematical, Scientific and Statistical Applications Software
Session Coordinator: Kenneth V. Byram

STORET - The Status of the System
Speaker: Charles S. Conger — Page 9

Abstract: A history of the STORET system, from its beginnings more than ten years ago, to its status at the present time is presented. Functions of user assistance group and the programmer staffing demands of the system are discussed. Some idea of the diverse nature of data in the system, and the volume of data it contains, along with plans for the future are presented.

A Method for Improving User Access to STORET
Speaker: Michael Friedland — Page 37

Abstract: The need to manipulate data stored in the STORET data base imposes a relatively complex programming problem on the user. This author has developed a FORTRAN subroutine which performs the housekeeping involved in making STORET data available in card-image form to the relatively unsophisticated user; that is, the user with limited WYLBUR and STORET vocabularies, and essentially no JCL or programming knowledge.

Two Data Storage Programs for STORET
Speaker: Kenneth V. Byram — Page 46

Abstract: One of the difficulties in preparing a large batch of STORET data for input is the difficulty of proofreading it before it goes into the system. The DIP input format, for example, is easy to prepare but difficult to check. This program tabulates an input data set in any format before data storage and also adds the ability to check values against upper and lower limits.

The EPA Scientific Applications Software Study
Speaker: Elijah Poole — Page 49

Abstract: EPA uses several computer facilities encompassing many intermediate and large-scale digital computers. The smaller systems are generally special purpose computers that service a single laboratory, while the large systems have remote terminals serving a large user community. Recent EPA computing system requirements studies have excluded "purely scientific" systems, and the scientific applications study now getting underway is designed to help in determining and fulfilling the Agency's needs in this area.

Statistical Analysis System (SAS)
Speaker: Richard Johnson — No Paper

Abstract: SAS is developed and updated by North Carolina State University, Department of Statistics. Released in 1969 for public consumption, there are now some 30,000 users of the system in approximately 170 installations and five foreign countries. Some of the reasons for its success and growth are discussed.

BIOMED, SAS, and Other Tools of the Statistician
Speaker: Robert Kinnison Page 50

Abstract: A comparision of the capabilities of the standard BIOMED package and of SAS is made from a user's point of view. The new P-Series BIOMED package, which is as yet unreleased, is introduced. Some available analytical routines, such as Goodman's nonparametric anova, chemical kinetic, and simulation packages are described.

MSP (Model Specification Program)–An Easy-to-Use Linear Statistical Analysis System
Speaker: Larry Male Page 56

Abstract: Most linear statistical programs severly limit a user to a fixed set of statistical models. Analysis of variance programs, for example, assumes a fixed model for estimating parameters and tests a fixed set of hypotheses. The user is unable to specify his own hypotheses, which may be of most interest. Although he can use "general linear model" programs, they require input of a cumbersome coefficient matrix with an entry for each experimental observation. The statistical analysis program described here allows a user to specify easily any number of models that he may want to analyze.

Concepts Verification Testing of the ENVIR System
Speaker: John Scotton Page 63

Abstract: The Data and Information Research Division, Office of Monitoring Systems, ORD contracted for six months with the Gulf Universities Research Consortium (GURC). The objective was to demonstrate and evaluate the application of the Environmental Information Retrieval (ENVIR) System to the collection, sorting, correlative analysis and interpretation of selected categories of EPA data.

This presentation discusses an evaluation relative to a concept verification testing of the ENVIR System on water, air and pesticides data bases. It also briefly discusses the comprehensive technical management information system known as Environment Dependant Management Process Automation and Simulation (EDMPAS) of which ENVIR is a primary model.

Interactive Computing - A Case History
Speaker: Richard Feldmann No Paper

Abstract: Time Sharing computer systems, as distinct from CRJE systems such as WYLBUR, have been available for some time, and in many applications have been enthusiastically utilized. This discussion compares the three modes of computing (time sharing, batch, remote job entry) pointing out the advantages and disadvantages of each mode for various applications (program development, graphics, number crunching, and information retrieval). The material presented reflects the authors experience with the NIH PDP-10.

SEAS - The Strategic Environmental Assessment System
Speaker: Edward R. Williams Page 69

Abstract: The Strategic Environmental Assessment System (SEAS) is a computer-based forecast and impact analysis technique that provides a national economic forecast which associates industrially related environmental residuals plus environmental treatment effects and costs. Additionally, it allows these factors to be distributed at a Federal region and state basis as well as forecasting environmental residuals associated with transportation and space conditioning. The model, operating in prototype form in the OSI computer, is being applied in analysis for Cost of Clean Air, Economics of Clean Water, Energy Conservation Analysis, Support of the Industrial Sludge Task Force, the Council for Environmental Quality, the National Commission for Water Quality and National Institues of Health.

Session II
October 3, 1974

Session Title: Applications of Minicomputers
Session Coordinator: Bill Budde

A Computer Controlled Data Acquisition and Retrieval System for Use in Air Monitoring Programs
Speaker: Marvin Hertz Page 88

Abstract: The heart of the CHAMP network is an automated data acquisition system employing intelligent controllers at both a central site and a network of remote monitoring stations. A minicomputer at each of the local stations controls and acquires data from aerometric instrumentation and transmits the data by phone lines to a central computer facility where the values are recorded and validated for use in health effects studies. Retrieval and processing of a large data base such as that acquired from the CHAMP network require a specialized approach. Total system capabilities are realized using a dual processor controller in the central facility operating in a foreground/background mode; semi-real time acquisition via telecommunications system and data recording are handled in foreground, and processing, validity checking, tabulation and reduction are accomplished as background tasks. A functional description of the system is given, including a data flow diagram and a discussion of the machine validation criteria. Emphasis is placed on the software considerations which make the CHAMP system unique.

An Aquatic Ecosystem Simulator and VS/8
Speaker: David M. Cline Page 96

Abstract: Aquatic Ecosystem Simulator (AEcoS), a one million dollars computer-controlled research facility used for physical simulation of freshwater ecosystems, is described with major emphasis on its computer system. The computer system consists of a PDP-8/E minicomputer and a real-time virtual operating system (VS/8) which is being developed at the Southeast Environmental Research Laboratory.

An On-Line PDP-12 Spriometry Package That Supports Both Onsite and Remote Stations
Speaker: Sam Bryan Page 99

Abstract: A spirometer (literally a "breath meter") is a device that can measure rate of air flow. When a person breathes through a tube connected to a spirometer, it can be used to measure how fast his lungs move air and how much air is moved. An electronic spirometer can be used in conjunction with a computer to provide a large array of measurements that relate to lung function and can provide them instantaneously. This paper describes such a system that is based around a PDP-12. This particular system can support an electronic spirometer that is connected directly to the computer or one that is located at a remote site that is connected through an analog telecommunications link. A graphics terminal at the remote site gives the attending technician feedback on the computer processing. A 1200 baud line is used for this digital telecommunications link.

An On-Line Real-Time Multi-User Laboratory Automation System
Speaker: Bill Budde Page 104

Abstract: Small to medium scale time sharing chemical laboratory automation systems were first developed during the late 1960's. The Methods Development and Quality Assurance Research Laboratory (MDQARL) and NERC-Cincinnati Computer Systems and Services Division are working with the Lawrence Livermore Laboratory (LLL) of the Atomic Energy Commission (AEC) to develop a time-sharing laboratory automation system with several novel features. The principal goal of the project is to automate presently owned chemical analysis instrumentation that is currently widely used at MDQARL and a number of other EPA laboratories for water quality parameter measurements. The significant feature is to do the automation in a fashion that will permit transfer of this technology to other EPA laboratories at a very significant cost and time savings. Another feature is the open-ended design that will permit the attachment of nonwater parameter instrumentation to the same system. These features will be substantially supported by the choice as the principal programming language of an extended version of BASIC. One other

NERC-Cincinnati research laboratory, an Office of Enforcement and General Counsel Field Center, and the Region V Surveillance and Analysis Laboratory have chosen to participate in this development. The salient features of this system are described with the exception of the sample file control program which is presented later in the workshop.

Computerized Gas Chromatography-Mass Spectrometry
Speaker: Craig Shew Page 114

Abstract: One of the most widely used automated experiments in EPA is the computerized gas chromatography-mass spectrometry system. About 25 complete systems valued at nearly three million dollars are available in a variety of laboratories for the identification and measurement of organic pollutants in environmental samples. Most of these systems emply 8K PDP-8/E or M minicomputers with a real-time disk operating system. Several systems use Varian 620 and Hewlett-Packard 2100 minicomputer systems. Cathode ray tube (CRT) displays are used on many of the systems for alphanumeric and graphic display at 1200 characters/second and user interaction. Programs have been developed for CPU-CPU communications with remote time-sharing systems. The PDP-8 based system which controls the spectrometer scan, acquires data, does data reduction, and outputs data are described.

Session III
October 3, 1974

Session Title: Applications of Interactive Graphics in Analysis of Environmental Monitoring Data
Session Coordinator: David M. Cline

Calcomp Software
Speaker: Theodore R. Harris Page 118

Abstract: Presentation is a discussion of EPA Computer Graphical capabilites. Basic information is given on Calcomp Plotting software and Tektronix CRT software.

Harvard Graphics Packages
Speaker: Curtis S. Lackey Page 132

Abstract: Presentation is a discussion of the Harvard Graphics Packages. The status of the system as well as applications are presented.

EPA's Regional Air Pollution Study (RAPS)
Speaker: Robert Jurgens Page 148

Abstract: Presentation is a paper entitled "Regional Air Pollution Study Graphics System." The development and current status of EPA's Regional Air Pollution Study (RAPS) and the methodology used from collecting and accessing the RAPS data bank are briefly reviewed. Computer graphics provide one of the primary media for the presentation of RAPS data. Justification of graphics for RAPS and hardware/software support at EPA, RTP are discussed. The graphics system which is being developed under contract is explained including its philosophy, hierarchal structure and classes displays.

Capabilities of Tektronix Software
Speaker: David M. Cline Page 156

Abstract: Presentation is a brief tutorial on the use of the Tektronix software as installed on the OSI computer facility as well as a PDP-8/E minicomputer. The Terminal Control System, the Advanced Graphics II System, and the Calcomp Preview Routines are discussed.

Session IV
October 3, 1974

Session Title: Laboratory Data Management
Session Coordinator: Wayne Ott

Automated Laboratory Management System in Region V
Speaker: David Rockwell — Page 159

Abstract: The automated handling of point source compliance monitoring data is presented for Region V Surveillance and Analysis Division. The dominant characteristic of this system is the use of automatic processing from the conception of a field survey to the storage of data in STORET. The programs involved in the automatic processing are LABORATORY-LABEL-INTERACTIVE-SIDES and QUALITY CONTROL.

Automatic Data Processing and Regional Laboratory Quality Assurance
Speaker: Billy Fairless — Page 167

Abstract: The Region V Central Regional Laboratory (CRL) will perform analyses during the current fiscal year. Sample types range from the very pure surface water of Lake Superior to sewage sludge. Concentrations vary from subnanograms per liter to essentially pure compounds and analyses are routinely requested for over 200 different parameters. This variety of problems presents a formidable challenge to the chemist to maintain control of each of his experiments and the quality of the resulting data. Steps the CRL have taken to define and measure the variables needed to evaluate the quality of data reported for each of several parameters are discussed. Methods are suggested in which automatic data processing might be used to assist laboratory personnel.

Sample Tracking Data Management System
Speaker: Robert N. Snelling — Page 182

Abstract: A Sample Tracking Data Management System (STDMS) is under development at NERC-LV to provide management information relating to sample analysis status as well as result reporting. Approximately 3000 samples per month are received by the laboratory for stable and radiochemical analysis. With an average of three analyses per sample, this results in approximately 9000 sample-analysis combinations per month to track. The system utilized two files: a Master file containing sample identification, collection, and parametric data; and a Status file maintaining statistics related to analysis status. Completed analysis results added to the Master file serve to update the Status file. An exception reporting technique is utilized to access information on number of samples received, analysis requested, analysis completed, and analysis overdue. Status data may be retrieved for a given sample, analysis category, or program.

SIDES
Speaker: David R. Barrow — Page 194

Abstract: SIDES (STORET Input Data Editing System) was developed to facilitate the preparation of data for entry into STORET. SIDES offers the following features: an easy-to-use card format for entering data particularly from the standpoint of keypunching; a straightforward method of entering sample identification information; a comprehensive pre-STORET editing system to allow a user to catch keypunch and transcription errors prior to entering the data into STORET; a simplified method of correcting data already entered into STORET through SIDES. Several data reports may be run from SIDES intermediate disk files to aid in data checkout and in obtaining data printouts with statistics prior to entry into STORET.

Selection of Probability Models for Determining Quality Control Data Screening Range Limits
Speaker: Wayne Ott Page 208

Abstract: Environmental monitoring laboratories carrying out routine chemical analyses increasingly are using computers to process and store the data. If these laboratories process large quantities of data, errors may be introduced in data handling phase, and there is some risk that these errors may go undetected. Such values, which frequently result simply from keypunch mistakes, may be stored as "valid" data and may create serious problems when sample means or other statistics are calculated. The computer can be used, however, to screen the data automatically as they are entered to determine if any value of a given parameter is "unusual", or lies outside an "acceptable range." Such unusual values can be "flagged" and brought to the attention of laboratory personnel where errors can be corrected and new analyses performed if necessary.

The Environmental Protection Agency (EPA) is now undertaking a data analysis effort in which probability models are fit to historical water quality data and range limits calculated. These limits then are being built into a quality control data screening computer program which serves as an intergral part of the computerized Laboratory Data Management System (LDMS). The basis for selecting and applying these probability models is discussed in detail.

A Systematic Approach to Water Quality Trend Analysis
Speaker: Merlin Dipert Page 217

Abstract: In investigating water quality trends, compounding of errors can be avoided by separating the errors of measurement from the errors of fitting. The fundamental data for consideration are the data points (individual points, replicated means, periodic means or some similar value and their calculated errors).

For the fitting procedure consider the general *n* dimensional mathematical model:

$$y=L(t+F(t)+G(x)+S$$

where $L(t)$ is the linear trend line used for the model, $F(t)$ is the nonlinear component of time, $G(x)$ is a function of the nonlinear components and S is a vector of seasonal components in the Box-Jenkins sense. Least squares estimates of the parameters are calculated and confidence limits in the quality control sense are drawn. The general model is relatively easy to fit since there is only one nonlinear function.

A computer program has been developed to retrieve the data from STORET and calculate the points and their errors, estimate the parameters of the curve in the least-squares sense, and plot the curves and data. The program is constructed so that the user can set his model or change it in one program statement. The user can fix any parameter desired by a control card. To date water temperature, nitrates, and phosphates have been fitted to the general harmonic model:

$$y=A+Bx+CSin\,(D+Ex)/EXP(Fx)$$

The nonsignificant parameters were set to zero the period fixed at one year. This article is a presentation of a cost-effective procedure for water quality analysis that is available as an analysis tool for qualified users.

Sample File Control for the Automated Laboratory
Speaker: Roman Bystroff Page 219

Abstract: Sample file control is to be an integral part of the automated laboratory minicomputer system which currently is being implemented for several EPA Labs as part of the ALMS program. Presentation discusses some aspects of the implementation and the philosophy of using a conservative systems approach.

A Data Reduction System for an Automatic Colorimeter
Speaker: Kenneth V. Byram

Page 226

Abstract: The System was designed to process results from a technicon autoanalyzer on a medium-size computer. The system creates sample and standard lists. The analyst acquires data rerun at higher dilutions. The system incorporates quality control and allows considerable flexibility in handling irregularities.

Session V
October 4, 1974

Session Title: General
Session Coordinator: Denise Swink

A Chemical Information System
Speaker: Stephen R. Heller

Page 234

Abstract: This presentation deals with those parts of a chemical information system which could be used by Agency Labs in their routine analysis work. The areas covered include literature, structure and hard data.

Mass Spectral Search Systems
Speaker: John McGuire

Page 236

Abstract: Setting standards and monitoring effluent streams requires a sensitive analytical tool for precise qualitative analysis. Computerized gas chromatography/mass spectrometry provides this tool for organic pollutants, but generates large amounts of data. Various approaches exist to simplify interpretation of these data. The EPA is developing three complementary systems to provide rapid, computer-controlled identifications from the data. One of these systems is now available to the public; the others will be in the near future. Presentation discusses the systems and their use in suppor of the EPA programs.

Bowne Time Sharing (WORD/ONE)
Speaker: Catherine Tittle

Page 254

Abstract: Ms. Catherine Tittle is Bowne Time Sharing's Washington Manager for Customer Service. She discusses the use of BTS's computerized text-processing service, Word/One, within the Environmental Protection Agency. Word/One has demonstrated effectiveness within the research community, and EPA specifically, where it has been used in the preparation of documents such as EXPRO, OR&D Program Planning & Reporting Manual, and Methods for Chemical Analysis of Water and Wastes (1974). Ms. Tittle discusses how Word/One may reduce the time and cost required for document preparation while simultaneously increasing production.

Introduction to the UNIVAC 1110 at Research Triangle Park-Capabilities
Speaker: T. L. Rogers

Page 257

Abstract: This is a presentation of the hardware configuration and capabilities of the UNIVAC 1110 System installed at the Research Triangle Computing Center. Included is a discussion of the processors, storage, and peripheral subsystems as installed. A description of the communications support including line types, number of lines, and terminals is also presented.

RTCC Software and Accessibility
Speaker: Maureen Johnson

Page 265

Abstract: This presentation is a brief description of the various software packages available on the UNIVAC 1110 at RTCC. This does not include detailed utilization techniques, but is an overview of available software resources. Information on how to become a registered RTCC user and obtain pertinent documentation is also included.

Technical and Environmental Information System
Speaker: Donald L. Worley

Page 271

Abstract: TENIS is an Information Storage and Retrieval System for the EPA UNIVAC 1110 which is being used to support the Air Pollution Technical Information Center. TENIS was custom-designed to provide not only low-cost support to APTIC but also to provide the complete range of services. The system is written in COBOL and was developed by the DSD staff at Research Triangle Park, N.C. Presentation discusses the system and its use to support APTIC.

The Conversion of CHESS and Other Systems
Speaker: Gene R. Lowrimore

Page 282

Abstract: This presentation is a description of experiences of one large user with the conversion effort from the IBM 360/50 to the UNIVAC 1110 system at RTP. The changes precipitated by differences in hardware and software architecture are emphasized. The things which worked well in the conversion and those that did not are presented from one user's viewpoint. Recommendations are made for how to approach a conversion effort of this type. Plans to change the systems to take advantage of new capabilities which the UNIVAC 1110 offers are discussed.

Present State of EDP Policy and Development in EPA
Speaker: Theodore R. Harris

Page 289

Abstract: Presentation is an informal discussion of current items which Management Information and Data Systems Division (MIDSD) are currently undertaking. These items include the EPA Feasibility Study Order (2800), the EPA Manual, the Standard Data Communications Terminal Procurement, the Interim EDP Resources Procurement and the Feasibility Study for the EPA Data Communications Network. The Whitten Committee Report and its effect on the Agency are presented. The EDP Plan is discussed under this topic.

Status of the Washington Computer Center Procurement
Speaker: Denise Swink

Page 296

Abstract: The current status of the development, approval, and implementation of the consolidated EPA computer resource (WCC) is discussed.

STORET - THE STATUS OF THE SYSTEM

By Charles S. Conger

The STORET concept, using magnetic tape as the storage medium, was conceived in 1962 to support the 120-station water quality network. STORET, an acronym for STOrage and RETrieval, is a computerized data base utilized by the Environmental Protection Agency (EPA), state and local pollution control agencies to define, record, and monitor the cause/effect relationship of water pollution. The system consists of several files, the principal files being the Water Quality File (WQF) and the General Point Source File (GPSF). Five basic areas are important in consideration of these two files (Figure 1).

The first area on the effects side, is the water use categories, which the states have defined by stream segment. The second area consists of sets of criteria or standards to describe these uses; the standards were established with federal approval by the states in order to numerically and quantitatively describe the uses. The third area and largest file in the STORET system is the WQF which describes the state of art of the river today, yesterday, a week ago, a year ago, and so forth. If the quality is known, it can be compared with the criteria and standards to determine whether or not the desired use is being produced. The above three areas describe the effects of water pollution.

The next consideration is the cause of these effects. In the fourth area, we attempt to define these causes by compiling a discharge inventory of point sources of pollution in order to determine what discharges are affecting water quality. The next question concerns the fifth area: What can be done to either upgrade or eliminate the discharges that are causing problems? Historically an action file of implementation plans has been maintained; this file describes when and how to upgrade or eliminate the problem discharges. These cause and action files have now been combined into the GPSF. The GPSF will consist minimally of 12 data collection efforts.

In order to integrate all of these files and to make the STORET system meaningful, five identifiers are associated with each piece of data. These five identifiers are action, use, parameter, location and time. A complete list of the pollution abatement actions available to EPA is contained in Figure 2. Typically these actions may be the granting of money for construction or court action resulting from Enforcement of the Permit System. The second identifier, use, is described in terms of Standard Industrial Classification (SIC) codes which are also listed in Figure 2. The third identifier, parameter, is grouped by the following categories: physical, chemical, biological, microbiological, waste treatment and economic. For any of this data to be of value, the user must know the fourth identifier, location, which is categorized as follows: political, hydrological, and geographical. Finally, the user must have an indication of the time frame of collection or occurrence.

Assuming that most of the readers are familiar with the mechanics of the system, let us examine the growth and content of the WQF. One of the more interesting growth indicators is a plot of the number of stations in the system versus time (Figure 3), which has some interesting inflections. Point A indicates where the system started utilizing low-speed remote terminals and where EPA initiated efforts to bring states on line. Point B occurs at the end of the USTS to BCS conversion, while Points C and D cover the first six months of operation at OSI. Point A also has some additional significance since it was at this point that the support staff of the system grew beyond two people. At point B the user assistance group was created. Presently there are four people in the programming support group which began utilizing software contracts to develop new software to support the GPSF in 1970. It is believed that the support group should be twice its size and should contain at least two junior people to accomplish such things as library maintenance, backup functions, and so forth. The user assistance group now has four people in it and needs twice that many to support its functions, including maintenance of all user routines. Lately, the user group has been performing extra duties in terms of developing input routines and testing contractor-supplied GPSF software. Never in its history has the user assistance group been in the position to initiate contact with the user community as opposed to reacting to user problems. EPA has recently awarded a contract to maintain and disseminate the STORET Manual. In the last few months, data element definition and contact with new users at the headquarters level have been handled by another group within the Monitoring and Data Support Division, which at least has enough manpower to complete the job adequately. Since the move to OSI, there has been an attempt to shunt as many of the communications and central systems related problems as possible

to the OSI project officer and the OSI customer support group. However, users do not make a distinction between the hardware and software. If a phone line is busy or the central system is down, STORET is not working as far as the user is concerned. Ideally, given an adequate computer utility, the staffing pattern to support STORET should consist of a user assistance group of eight to ten people, a systems/programming group of eight to ten people, and an input/data element definition group of eight to ten people plus the continuation of contract support for the STORET Manual. A real asset would be a group of two people to keep track of utilization, budget problems, justification and defense of the system, and so forth. During the last six years the system has been in a continuous state of audit; and in fact, at the present time, there are at least two studies occurring simultaneously. To respond to questions and requests from audit teams, resources from the programming staff and from user assistance must be drawn upon; a management staff function would eliminate the necessity for this.

Growth and utilization statistics follow. Figure 4 indicates the level of activity against the WQF during a typical week. Figure 5 summarizes the data growth of the WQF. Figures 6 and 7 show a historical level of activity against the WQF. The GPSF has not been in existence long enough to show historical trends, but the present content and level of activity is shown in Figure 8.

New developments include the publication of a STORET System Directory. This directory is designed to aid the new user in finding stations with enough and/or the particular data of interest. Figure 9 is an example indicating the number of parameters at particular stations. Figure 10 may be utilized to pinpoint the location of a particular station on USGS maps.

Quite a bit of effort has been placed on developing graphical display of data from the WQF. The Multiple Station Plot (MSP) package has been available for over two years, but its new features are probably not adequately documented for the unsophisticated user. Figures 11 through 16 contain examples of features taken from MSP. Another plot capability that is certainly not utilized a great deal is PGM=LOC. A typical example can be examined in Figure 17. Figure 18 shows a plot produced from a CALCOMP subroutine. In terms of plans for the future, a WQF preedit program is being designed to provide at least a minimal amount of quality control on input data.

Assuming that EPA will have access again someday to a system that can support interactive graphics, a pilot project will be implemented under an R & D grant completed two years ago with the state of Michigan (see Figures 19 thru 28). Figures 29 and 30 are examples of graphical output from the AUTOMAP subsystem and the Municipal Waste Inventory, respectively.

The following story illustrates the importance of considering both how you use the computer as well as what you use it for. Once upon a time in a house in the woods lived three bears: Daddy Bear, Mummy Bear and Baby Bear. Now Daddy Bear was called ACCOUNTANCY, and he made sure that the family got what it needed, not what it wanted. He was very fond of making tables, which he did by simply nailing five pieces of wood firmly together. However, he had a clever knack for making all four legs exactly the same length so that the Bears' tables hardly ever wobbled.

Mummy Bear was called STATISTICS and she took care of the cooking. Although the Bears ate only porridge, she knew their likes and dislikes and could usually give them the sort they liked. Of course it meant a lot of work keeping her recipe book up to date.

Baby Bear was called MATHEMATICS and he had a tin whistle. On it he could compose catchy little tunes that for some reason he called theorems. Although he really was not very reliable, one of his little masterpieces could sometimes help his parents over a bad patch.

One day when ACCOUNTANCY, STATISTICS and MATHEMATICS were away, presumably at a conference, who should come to the house but little GOLDILOCKS COMPUTER. Well of course she tried Daddy Bear's tables, and tasted Mummy Bear's porridge, and blew Baby Bear's whistle. She was surprised to find that although the tables were very plain, they did not wobble. The porridge in each plate tasted different even though, as far as she could see, the only food in the house was plain porridge; and all she could get out of Baby Bear's whistle was a faint squeak.

Anyway, the Bears returned home and found her there. Because she really was a charming young lady, they asked her to stay; and in the fullness of time, she felt she ought to do them a good turn.

She told Daddy Bear that she would help him design some really elegant tables. She asked Mummy Bear if it really would not be better to make a big pot of basic porridge and then let the family get what they wanted

from it. Mummy Bear was a bit doubtful about the ingredients of basic porridge, but Goldilocks said she would help sort it out. She also made it quite plain to Baby Bear that he ought to have a stock of tunes ready for bad times.

As it happened the Bears thought that these ideas were worth trying. Daddy Bear had always felt his tables could be improved, Mummy Bear was fed up with always changing her recipe books, and Baby Bear, to tell the truth, was a little guilty about being selfish. So they agreed to Goldilock's plans.

At first everything went well. Plans for splendid tables were drawn up. Several dozen recipes for basic porridge were prepared, and Baby Bear began daily music practice. However, the bears could not help noticing that they were all so busy that they had no time left for making tables, cooking porridge or composing tunes. But they still felt everything would be all right.

Then came the great day when they were ready to start their new life. Unfortunately Daddy Bear soon found that his new tables, which had at least a hundred parts, wobbled when he could manage to get them together. Mummy Bear realized that the only way to settle the recipe for basic porridge was to carry out a survey of Bears' attitudes towards porridge. She knew, however, that after some weeks of very thin porridge, their attitude to it was distinctly hostile. It was clear that she could not complete her survey until she had made some good basic porridge, and she could not find out how to make good basic porridge without a survey. Baby Bear said he could not get any inspiration under pressure and locked himself in his bedroom.

The next day the Bears all asked Goldilocks Computer to help them out of their trouble. However, Goldilocks Computer was very busy looking into the possibilities for making automatic reclining chairs, cooking coq-au-vin and playing alto saxophone. So she said, "I really know nothing about the actual job of making tables, cooking porridge and playing tin whistles. But I have given you all the information you need, so why don't you get on with it?" Now Bears, although nice enough characters, are only Bears and they have very little sense of humor, especially where tables, porridge and music are concerned.

So Daddy Bear told Goldilocks Computer to stop interfering with his tables, and he showed her his long black claws just to make sure she understood. Mummy Bear got out her old recipe book, threw away the mess of basic porridge, and let Goldilocks see just the tips of her sharp claws. When little Mathematics Bear stopped his practice and showed Goldilocks his claws, she was surprised to see how long and sharp they were and went to the corner and cried. Soon afterwards her beautiful golden hair turned grey. Because she could not make tables, cook porridge or play tin whistles, there was nothing left for her to do but the washing up. And if you have ever washed up after a party of porridge eaters, you know how she felt.

Why use a moral story at a serious conference? Well, the first reason is to draw the moral. To stress it, perhaps even to labor the point, the moral is: Ask to see bears' claws before telling them what their problems are. That moral is commended to all the computer men, analysts and so on present. Engineers and managers are also urged to keep it in mind.

The lesson is that it matters at least as much how you use the computer as what you use it for. The conference was not intended to promote computer use. What it has done is to bring together many diverse interests so that each of them can hear what the others have to say about how they do things.

STORET

WATER QUALITY EFFECT

USE	CRITERIA OR STANDARDS	WATER QUALITY

GENERAL POINT SOURCE

CAUSE	ACTION
DISCHARGE INVENTORY	IMPLEMEN-TATION PLANS

MUNICIPAL WASTE INVENTORY
INDUSTRIAL WASTE INVENTORY
- PERMIT FILES
  - MICS
  - SHORT FORMS
  - LONG FORMS
  - SELF MONITORING
  - RAPP
- VOLUNTARY

CONTRACT AWARDS
CITY MASTER FILE
POLLUTION CAUSED FISH KILLS
FEED LOT/FISH FARM (AGRICULTURE)
MINE DRAINAGE
DEEP WELL INJECTION SURVEY
MUNICIPAL DRINKING WATER SUPPLY
NEEDS SURVEY
OCEAN DUMPING

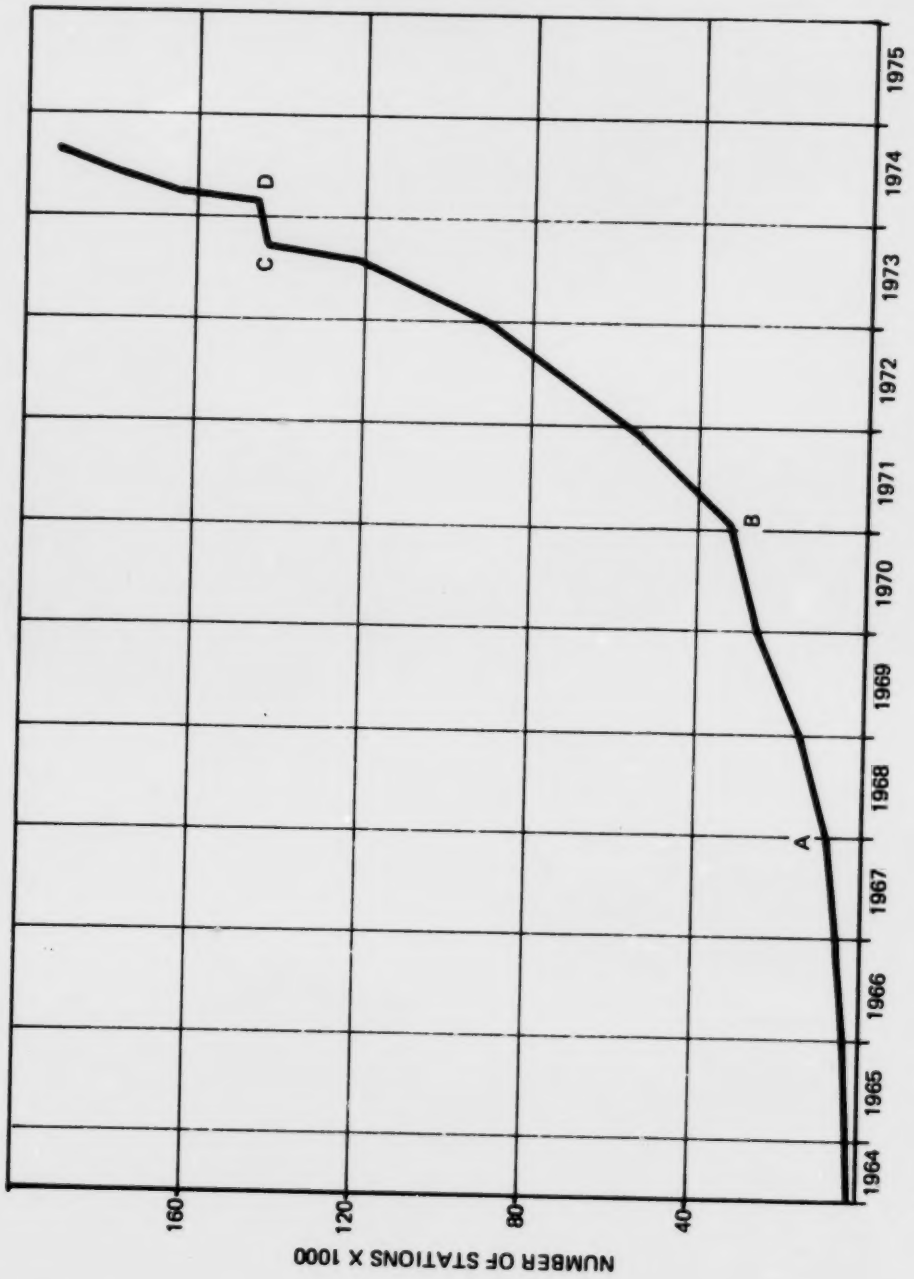
Figure 1

STORET

Action	Use	Parameter	Location	Time
(Federal, State, Local) Effluent standard criteria established Effluent quality measurement Voluntary industrial waste inventory RAPP FPC survey Influent quality measurement Voluntary industrial waste inventory RAPP FPC survey Enforcement 180 day notice Conference Court order Discharge permit Waste treatment facility construction Preliminary plan submission Preliminary plan approval Final plan submission Final plan approval Financing complete Site acquired Grant award Contract award Begin construction End construction and start discharge Operational level attained Waste treatment facility operation Training Technological advancement WQ measurement WQ standards established Use stopped	SIC Agriculture, forestry fisheries Mining Contract construction Manufacturing Transportation, communication, electric gas, and sanitary services Wholesale and retail trade Finance, insurance, and real estate Services Government Organization Plant Discharge Life support	Physical Chemical Biological Microbiological Waste treatment facilities (e. g.) Primary Intermediate BOD Removal Secondary BOD Removal Tertiary Phosphate Removal Disinfection Sludge processing Sludge disposal Outfall Deep ocean outfall Sanitary intercepter sewer Combined intercepter sewer Pumping station Force main Combined sewer overflow Economic Men Dollars Numeric	State County Cong. Dist. City RMI Lat. /Long.	Planned (year, month, day hour, minute, second) Executed (year, month, day hour, minute, second) Map observation time

Figure 2

Figure 3


NUMBER OF STATIONS X 1000
160
120
80
40
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
A
B
C
D

1974

w/o [illegible]

[illegible] 21[illegible]

YEAR	ACTIVE STATIONS	TOTAL VISITS	TOTAL OBSERVATIONS		[illegible]	[illegible]
<1930	137	10441	28855		[illegible]	
1930	68	1006	3498	[illegible]	[illegible]	[illegible]
1931	40	933	2337	[illegible]	[illegible]	[illegible]
1932	631	2134	13333	[illegible]	[illegible]	[illegible]
1933	261	1552	7781	[illegible]	[illegible]	[illegible]
1934	145	1439	7961	[illegible]	[illegible]	[illegible]
1935	159	2082	9321	[illegible]	[illegible]	[illegible]
1936	167	1461	7157	[illegible]	[illegible]	[illegible]
1937	239	1283	10669	[illegible]	[illegible]	[illegible]
1938	321	2268	13504	[illegible]	[illegible]	[illegible]
1939	241	1764	10086	[illegible]	[illegible]	[illegible]
1940	345	2706	11882	[illegible]	[illegible]	[illegible]
1941	893	3305	20652	[illegible]	[illegible]	[illegible]
1942	419	2646	11472	[illegible]	[illegible]	[illegible]
1943	522	2780	13370	[illegible]	[illegible]	[illegible]
1944	545	3071	17408	[illegible]	[illegible]	[illegible]
1945	662	4052	27857	[illegible]	[illegible]	[illegible]
1946	1124	6647	50829	[illegible]	[illegible]	[illegible]
1947	1205	10920	81359	[illegible]	[illegible]	[illegible]
1948	944	6209	50893	[illegible]	[illegible]	[illegible]
1949	1205	7993	66109	[illegible]	[illegible]	[illegible]
1950	1630	9429	76684	[illegible]	[illegible]	[illegible]
1951	2452	14122	102191	[illegible]	[illegible]	[illegible]
1952	2570	14840	107066	[illegible]	[illegible]	[illegible]
1953	2778	17944	126813	[illegible]	[illegible]	[illegible]
1954	3622	22164	155095	[illegible]	[illegible]	[illegible]
1955	3150	18633	140587	[illegible]	[illegible]	[illegible]
1956	3409	20710	154253	[illegible]	[illegible]	[illegible]
1957	3320	24183	172925	[illegible]	[illegible]	[illegible]
1958	4391	32712	252782	[illegible]	[illegible]	[illegible]
1959	6361	39323	334276	[illegible]	[illegible]	[illegible]
1960	6249	60849	507005	[illegible]	[illegible]	[illegible]
1961	6581	82334	620475	[illegible]	[illegible]	[illegible]
1962	8505	94294	823740	[illegible]	[illegible]	[illegible]
1963	9803	101863	886792	[illegible]	[illegible]	[illegible]
1964	9446	105636	948960	[illegible]	[illegible]	[illegible]
1965	12750	147761	1258332	[illegible]	[illegible]	[illegible]
1966	15600	153823	1360422	[illegible]	[illegible]	[illegible]
1967	17716	176550	1561032	[illegible]	[illegible]	[illegible]
1968	18099	196372	1681105	[illegible]	[illegible]	[illegible]
1969	20443	215193	1847054	[illegible]	[illegible]	[illegible]
1970	24274	253912	2283319	[illegible]	[illegible]	[illegible]
1971	28759	227775	2542824	[illegible]	[illegible]	[illegible]
1972	35647	313548	2765153	[illegible]	[illegible]	[illegible]
1973	32027	268439	2475883	[illegible]	[illegible]	[illegible]
1974	7766	28324	252774	[illegible]	[illegible]	[illegible]
>1974	[illegible]	[illegible]	[illegible]			
TOTAL		2721931	23906632			

Figure 4

STORET SYSTEM

WATER QUALITY FILE

RUN DATE: 10 / 02 / 74

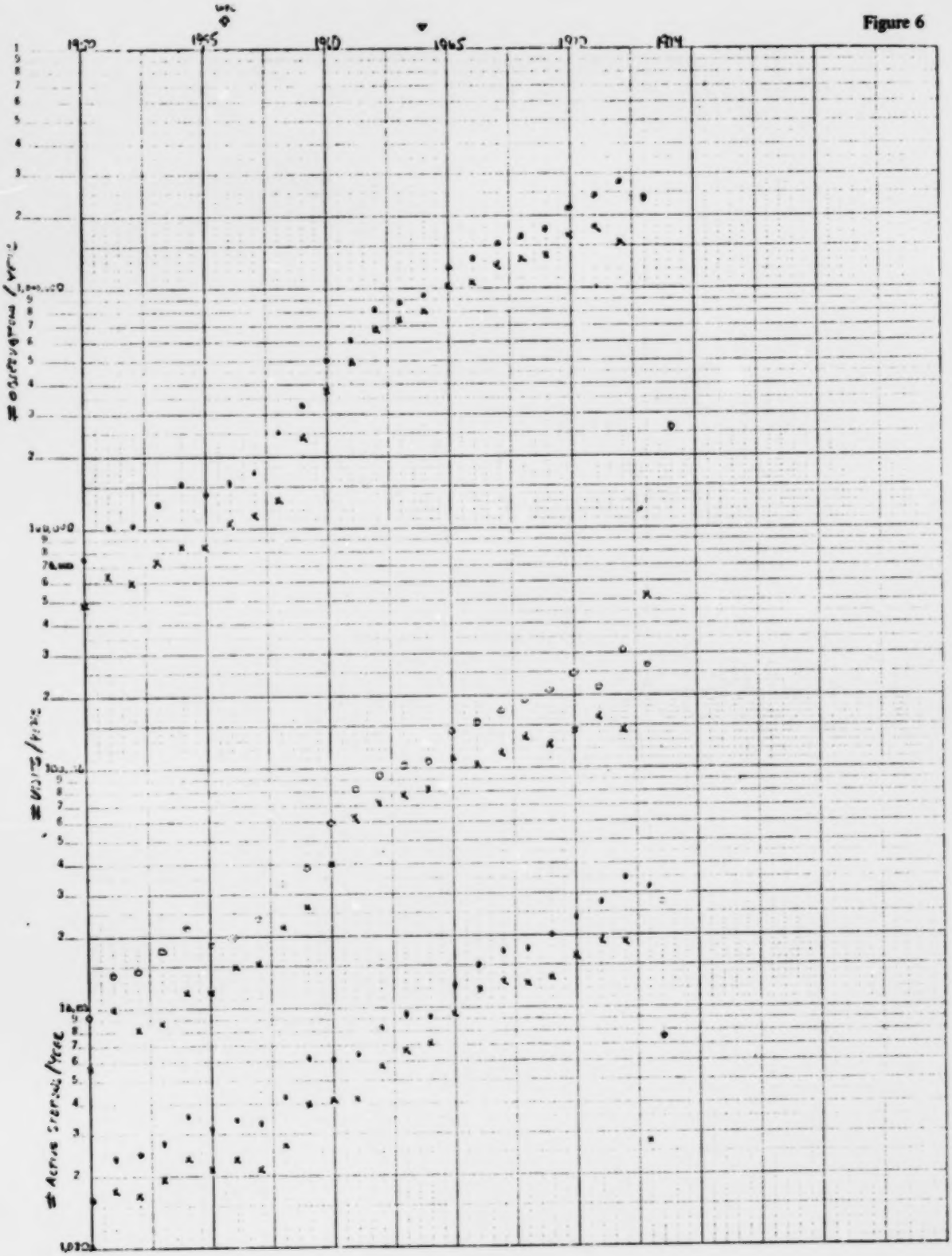
SUMMARY OF DATA STORAGE INPUT

FOR WEEK OF 07 / 26 / 74

USER	STATIONS NEW	STATIONS CHANGE	STATIONS DELETED	WATER QUALITY DATA STATIONS	NEW PARMS	DATA OBSVS	CHANGED PARMS	DATA OBSVS	DELETED STATIONS	PARMS	DATA OBSVS	DATE ENDING	DATE BEGIN
				1	0	0	0	0	0	0	0	0	999999
UNIVWASH	11	0	0	10	1	1	0	0	0	0	0	0	999999
11COEDED				96	36	4772	0	0	0	0	0	731211	730212
11FRALES	1	0	3	154	22	5806	0	0	1	0	0	740626	720505
11USAFA	1	0	0	3	3	3	0	0	0	0	0	740225	730228
11VEGA49	1	0	0	16	45	97	0	0	1	0	0	740711	730703
1110NET	22	0	0	90	17	741	0	0	181	0	0	740226	720213
1111H010	1	1	0	11	16	128	0	0	0	0	0	740701	620511
1111H030				72	26	480	0	0	0	0	0	740702	740417
11112300	3	0	0	2	3	3	0	0	0	0	0	0	999999
11115000	48	0	0	91	20	578	0	0	0	0	0	740701	680901
1113APWQ	1	0	0	8	4	6	0	0	0	0	0	671101	671005
1113CBDP				25	7	277	0	0	0	0	0	680319	680319
1113COOP	1	0	0	9	25	155	0	0	0	0	0	740713	740605
1113ELRF				21	7	287	0	0	0	0	0	720719	720718
1113FISH	1	0	0										
1113JAWQ	1	0	0	22	8	64	0	0	0	0	0	711104	711019
1113MONA	1	0	0	5	2	2	0	0	0	0	0	0	999999
1113SUST				29	4	1055	0	0	0	0	0	720623	710714
1113S000	1	0	0	2	4	4	0	0	0	0	0	681017	681017
1113S050	2	0	54	28	1	1	0	0	0	0	0	0	999999
1113WSWQ	1	0	0	6	8	12	0	0	0	0	0	710412	681119
1115G050	1	0	0	4	21	782	1	3	0	0	0	740531	730312
1116APCC	71	0	0	71	47	2821	0	0	0	0	0	740514	430522

Figure 5

Figure 6



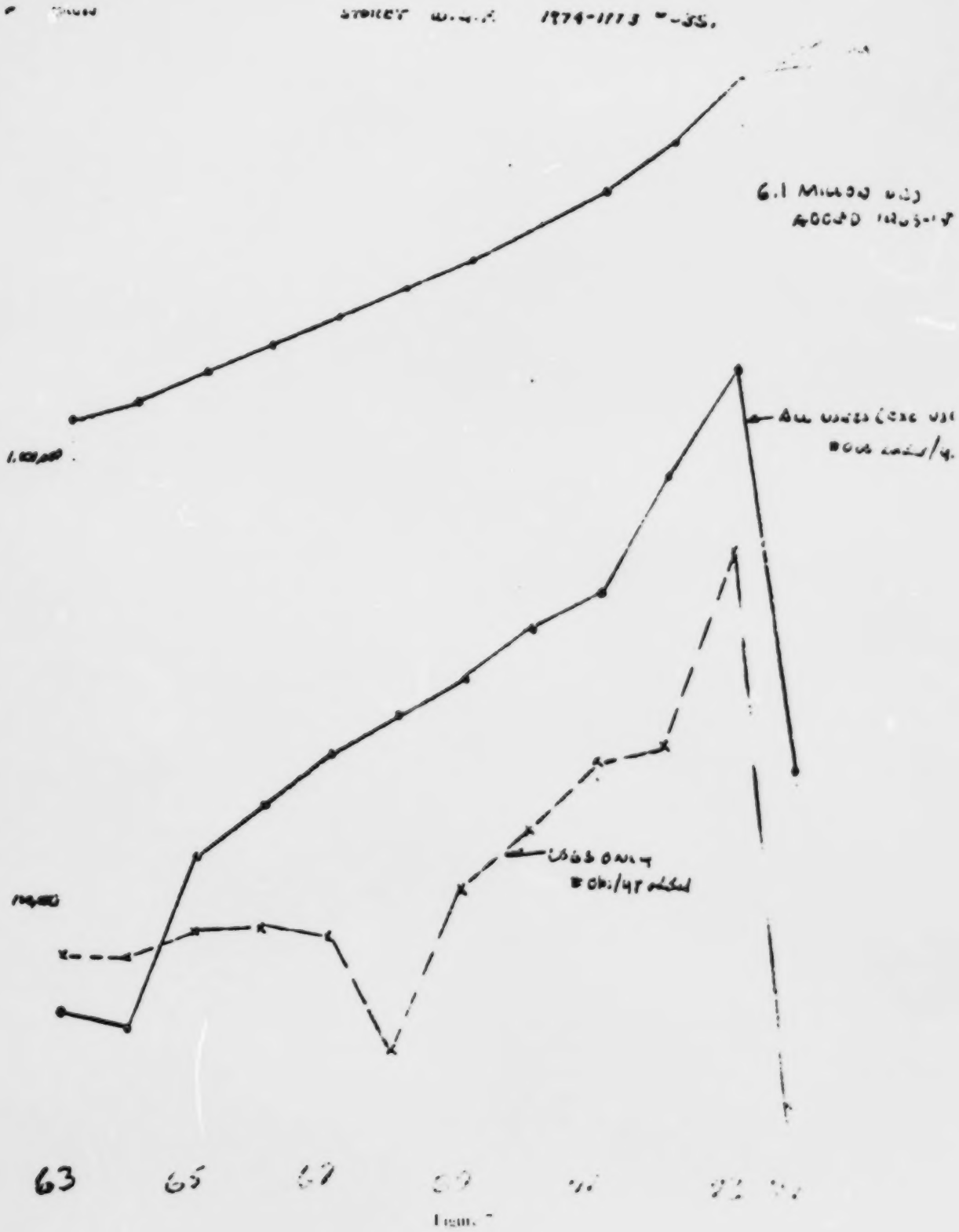


Figure 7

GPSF FILE CONTENTS

September 12, 1974

REGION	FACILITIES IN MICS: APPLICATIONS RECEIVED: PERCENT IN 9/12/74 VS. EXPECTED*	FACILITIES IN MICS: APPLICATIONS RECEIVED: PERCENT INCREASE[a] 8/6/74 TO 9/12/74	FACILITIES IN MICS: PERMITS ISSUED: PERCENT IN 9/12/74 VS. EXPECTED**	FACILITIES IN MICS: PERMITS ISSUED: PERCENT INCREASE[a] 8/6/74 TO 9/12/74	FACILITIES WITH SELF-MONITORING RECORDS: PERCENT IN 9/12/74 TO EXPECTED***	FACILITIES WITH SELF-MONITORING RECORDS: PERCENT INCREASE[a] 8/6/74 TO 9/12/74	FACILITIES WITH COMPLIANCE SCHEDULES: PERCENT IN 9/12/74 VS. EXPECTED****	FACILITIES WITH COMPLIANCE SCHEDULES: PERCENT INCREASE[a] 8/6/74 TO 9/12/74
1	96	34	49	15	3	3	70	27
2	100	0	88	1	57	49	6	-2
3	100	0	71	13	0	0	88	88
4	100	0	99	-1	49	42	62	8
5	91	1	73	-11	33	33	88	-11
6	100	3	6	-1	0	0	0	0
7	74	23	33	17	0	0	0.2	0.2
8	100	30	11	-1	7	0	0	0
9	75	0	41	-4	0.1	0.1	0.3	0
10	100	0	46	12	9	8	12	0
TOTAL	95	6	58	2	20	18	43	11

* EXPECTED NUMBER IS BASED ON THE QUARTERLY REPORT (AS OF 6/30/74) OF APPLICATIONS RECEIVED BY PERMIT ASSISTANCE & EVALUATION BRANCH, OEGC, OR TOTAL APPLICATIONS IN MICS, WHICHEVER IS GREATER.

** EXPECTED NUMBER IS BASED ON THE MONTHLY REPORT (AS OF 8/31/74) OF PERMITS ISSUED BY PERMIT ASSISTANCE & EVALUATION BRANCH, OEGC.

*** EXPECTED NUMBER IS THE SAME AS THE NUMBER OF PERMITS ISSUED (SEE ** ABOVE).

**** EXPECTED NUMBER IS DERIVED BY MULTIPLYING .85 TIMES THE NUMBER OF MAJOR PERMITS ISSUED AND ADDING .65 TIMES THE NUMBER OF MINOR PERMITS ISSUED (SEE ** ABOVE).

[a] NOTE:
PERCENT INCREASE IS OBTAINED BY SUBTRACTING PERCENT IN VERSUS EXPECTED ON 8/6/74 FROM PERCENT IN VERSUS EXPECTED 9/12/74.

Figure 8

PAGE 1
1ST EDITION
SEPT 1974
REGION 1

STATE 9

STORET SYSTEM DIRECTORY
FOR SELECTED STATIONS AND DATA
AS OF APRIL 1974

A	ST	C	STATION LOCATION	AGENCY	STATION	TYPE	EA	LA	NO	PH YSICAL	SO LIDS	NI TROGEN	PH OSPHORUS	DISSOLVED OXYGEN	OXYGEN DEMAND	RAINFALL	MEAN DISCHARGE	RAINFALL	RIVER FLOW	GEOLOGY	GROUND WATER LEVEL	PESTICIDES	FLOW	TEMPERATURE	LN
P	9		QUINEBAUG RIVER AT QUINEBAUG, CO	112WRD	01124000	02111204	60	62	2	39	17					97			17	77		1A	1A		1
H	9		HOLLENBECK R AT HUNTSVILLE, CT.	112WRD	01198990	77777777	72	73	2	4D		44	14	1A		C4	39			84	J2	18	1D		2
H	9		WAMGUM LAKE BK NEAR HUNTSVILLE	112WRD	01198851	77777777	72	73	2	4C				19		C3	39				J2	16	1C		3
P	9		WAMGUM LAKE BK NEAR SOUTH CANAAN	112WRD	01198880	77777777	72	73	2	4C				19		B3	39				G2	16	1C		4
H	9		HOUSATONIC R AT FALLS VILLAGE, C	112WRD	C1199003	77777777	72	73	2	4D		44	24	19		A4	39			84	J2	18	1D		5
P	9		FARMINGTON R AT TARIFFVILLE, CT.	112WRD	01189995	02111204	71	73	3	5G		55	28	1A		E6	3B			88	I2	1K	1K		6
H	9		FARMINGTON R AT PLAINVILLE, CT.	112WRD	C1190000	02111204	68	70	3	4F			17	17			17			77		17	17		7
P	9		CONNECTICUT RIVER AT SHELDON DAM	1110NET	080082	02111204	62	69	8	2Z	2E	38	48	1Z	1Z	CC	K4	1Y	A8	55	5Z		1Z	1Z		8
H	9		CONNECTICUT R AT THOMPSONVILLE C	112WRD	01184000	06111204	66	73	8	7H	4D	2C	2D	1N	1A	6C	B3	3H		1D	8C	B2	1N	1D	9
P	9		HELL MEADOW BK NR BROOKSVILLE, CT	112WRD	01205561	02111204	66	67	2	3A	1A	19				7A				19	7A			19	10
H	9		BURLINGTON BK NR BURLINGTON, CT.	112WRD	01188000	02111204	68	73	6	7C	26	18	29	1C		9C	3C				8C		1F	1F	11
P	9		CT.R-HOCKANUM R-600 FT BRW CONF.	1111PEG1	HR01	02111204	73	73	1	17	14	14	14	17	14	24			14					17	12
P	9		CT.R-HOCKANUM R-MAIN ST BRIDGE	1111PEG1	HR02	02111204	73	73	1	17	14	14	14	18	14	24			14					18	13
P	9		CT.R-HOCKANUM R-HILLSIDE ST BRG	1111PEG1	HR03	02111204	73	73	1	17	15	15	14	18	15	25			15					19	14
P	9		CT.R-HOCKANUM R-SCOTLAND RD BRG	1111PEG1	HR04	02111204	73	73	1	17	14	14	14	18	14	24			14					18	15
P	9		CT.R-HOCKANUM R-WALNUT ST BRG	1111PEG1	HR05	02111204	73	73	1	15	14	14	14	18	14	24			14					18	16
P	9		CT.R-HOCKANUM R-MID TURK BRG (W)	1111PEG1	HR06	02111204	73	73	1	16	14	14	14	18		24			14					18	17
P	9		CT.R-HOCKANUM R-UPSTR OF UNN BRD	1111PEG1	HR07	02111204	73	73	1	16	14	14	14	17	14	24			14					18	18
P	9		CT.R-HOCKANUM R-TOLLAND TURK BRG	1111PEG1	HR09	02111204	73	73	1	15	14	14	14	18	14	24			14					18	19
P	9		CT.R-HOCKANUM R-TOLLAND TURK BRG	1111PEG1	HR11	02111204	73	73	1	15	14	14	14	18	14	24			14					17	20
P	9		HOCKANUM R NR EAST HARTFORD, CT.	112WRD	01192500	02111204	60	69	3	3A	1G					7A					7M		1X		21
P	9		CT.R-HOCKANUM R-	1111PEG1	HR12	02111204	73	73	1	15	14	14	14	18	14	24			14					18	22
P	9		CT.R-HOCKANUM R-PART AL RD BRG	1111PEG1	HR13	02111204	73	73	1	15	14	14	14	18	14	24			14					18	23
P	9		CT.R-HOCKANUM-WRRG WOOD RD BRG	1111PEG1	HR14	02111204	73	73	1		14	14	14	16	14	24			14					16	24
P	9		BARLTVILLE LAKE	11EPALES	090401	02111202	72	72	1	26		26	26	14						16				14	25
P	9		HOUSATONIC R AT GAYLORDSVILLE CO	112WRD	01200500	06111204	60	68	2	2B	1B					8B				7B		1B	1B	18	26
P	9		BRYANT LAKE	11EPALES	090201	02111202	72	72	1	2A		2E	2A	18						1A				18	27
P	9		BANTAM LAKE	11EPALES	090202	02111202	72	72	1	26		26	26	14						16				14	28
P	9		SSEISHOOK LAKE	11EPALES	090101	02111202	72	72	1	38		25	25	16						15				16	29
P	9		SHEYUCKET RIVER NEAR WILLIMANTIC	112WRD	01122500	02111204	63	73	7	5E			1C	1C			39	3B		7C		1E	1E	30
P	9		HOUSATONIC LAKE LITCHFIELD	1111PEG1	LL05.55	02111202	73	73	1	24				17			22		17				18	31
P	9		HOUSATONIC LAKE LITCHFIELD	1111PEG1	LL055	02111202	73	73	1	24				17			22		17				18	32
P	9		HOUSATONIC LAKE LITCHFIELD	1111PEG1	LL065	02111202	73	73	1	24				17			22		17				18	33
P	9		HOUSATONIC LAKE LITCHFIELD	1111PEG1	LL075	02111202	73	73	1	26				15			22		16				17	34
P	9		HOUSATONIC STILL RIVER	1111PEG1	ST201	02111204	73	73	1	27	14	14		16	14		22		17	14			18	35
P	9		SOUTHBRIDGE, CONNECTICUT	1111PEG1	CR-J2.0	02111204	73	73	1	1G	14	14	14	18			24		18	14			18	36
P	9		SOUTHBRIDGE, CONN.	1111PEG1	CR-J2.5	02111204	73	73	1	1B	14	14	14	18			24		18	14			18	37
P	9		HANOVER POND	11EPALES	W09501	02111202	72	72	1	26		26	26	14						16				14	38
P	9		CHESTER, CONN.	1111PEG1	CR-J3.3	02111204	73	73	1	1B	14	14	14	19			24		18	14			18	39
P	9		WORTHEN, CONN.	1111PEG1	CR-J4.0	02111204	73	73	1	1B	14	14	14	18			24		18	14			18	40
P	9		WORTHEN, CONN.	1111PEG1	CR-J5.0	02111204	73	73	1	1B	14	14	14	18			24		18	14			18	41
P	9		CONNECTICUT R AT MIDDLE HADDAM,	112WRD	01193350	06111204	67	73	7	5C	2C		1C	1C			AC	39			8C		1D	42
P	9		SALMON R NR EAST HAMPTON, CT.	112WRD	01193500	02111204	60	73	8	5C			1C	1C			B9	39			7C		1C	1C	43
P	9		YANTIC R AT YANTIC, CT.	112WRD	01127500	02111204	68	73	6	6F			2D	1C			74	3C			8C		1E	1E	44
P	9		HOUSATONIC STILL RIVER	1111PEG1	ST202	02111204	73	73	1	27				17			22		18					18	45
P	9		HOUSATONIC STILL RIVER	1111PEG1	ST203	02111204	73	73	1	27	14	14		18	14		22		18	14				18	46
P	9		HOUSATONIC STILL RIVER	1111PEG1	ST203.5	02111204	73	73	1	27	14	14		18			22		18	14				17	47
P	9		HOUSATONIC STILL RIVER	1111PEG1	ST204	02111204	73	73	1	27	14			18			22		18	14				17	48
P	9		HOUSATONIC STILL RIVER	1111PEG1	ST205	02111204	73	73	1	27	14	14		18			22		18	14				18	49
P	9		HOUSATONIC STILL RIVER	1111PEG1	ST206	02111204	73	73	1	2B	15	15	14	19	14		22		19	15				19	50

Reproduced from best available copy.

Figure 9

PAGE 2
1ST EDITION
SEPT 1974
PERIOD 1

STREAM SYSTEM DIRECTORY
FOR SELECTED STATIONS AND DATA
AS OF APRIL 1974

STATE 9

LN	STATION	LATITUDE	LONGITUDE	MA SU	DWDCK	1:250,000 SERIES	USGS MAP NAMES P C R 7.5' OR 15' SERIES	EPA MAJOR AND MINOR BASINS
1	09000	42 1 20	71 57 22	02	G 63257	BOSTON	H 1 WEBSTER 7.5 NE05	
2	09105	41 56 55	73 19 18	02	M 73135	HARTFORD	A 6 SOUTH CANAAN 7.5 NE02	
3	09105	41 56 36	73 17 8	02	M 73146	HARTFORD	A 6 SOUTH CANAAN 7.5 NE02	
4	09005	41 57 42	73 18 52	02	M 73149	HARTFORD	A 6 SOUTH CANAAN 7.5 NE02	
5	09005	41 57 26	73 22 11	02	K 73150	HARTFORD	A 6 SOUTH CANAAN 7.5 NE02	
6	09003	41 54 30	72 45 40	02	K 73144	HARTFORD	A10 TARIFFVILLE 7.5 NE04	
7	09003	41 54 41	72 41 16	02	K 63530	HARTFORD	A11 WINDSOR LOCKS 7.5 NE04	
8	09103	41 59 25	72 36 21			HARTFORD	A12 ROAR BROOK 7.5 NE04	
9	09103	41 59 14	72 36 21	02	K 63528	HARTFORD	A12 ROAR BROOK 7.5 NE04	
10	09105	41 51 60	73 10 4			HARTFORD	B 7 WEST TORRINGTON 7.5 NE02	
11	09103	41 47 10	72 57 55	02	K 63529	HARTFORD	B 9	
12	09103	41 45 20	72 35 13			HARTFORD	B11 HARTFORD NORTH 7.5 NE04	
13	09103	41 45 41	72 38 45			HARTFORD	B11 HARTFORD NORTH 7.5 NE04	
14	09103	41 46 32	72 37 14			HARTFORD	B12 MANCHESTER 7.5 NE04	
15	09103	41 46 33	72 36 20			HARTFORD	B12 MANCHESTER 7.5 NE04	
16	09103	41 46 55	72 35 32			HARTFORD	B12 MANCHESTER 7.5 NE04	
17	09103	41 46 47	72 34 2			HARTFORD	B12 MANCHESTER 7.5 NE04	
18	09103	41 47 59	72 31 42			HARTFORD	B12 MANCHESTER 7.5 NE04	
19	09103	41 48 38	72 31 6			HARTFORD	B12 MANCHESTER 7.5 NE04	
20	09113	41 49 9	72 30 8			HARTFORD	B12 MANCHESTER 7.5 NE04	
21	09103	41 45 36	72 35 16			HARTFORD	B12 MANCHESTER 7.5 NE04	
22	09113	41 49 57	72 29 50			HARTFORD	B13 ROCKVILLE 7.5 NE04	
23	09113	41 50 60	72 29 20			HARTFORD	B13 ROCKVILLE 7.5 NE04	
24	09113	41 51 56	72 29 13			HARTFORD	B13 ROCKVILLE 7.5 NE04	
25	09030	41 46 60	72 16 60			HARTFORD	B14 SOUTH COVENTRY 7.5 NE05	
26	09100	41 39 11	73 29 25			HARTFORD	C 5 KENT 7.5 NE02	
27	09100	41 41 60	73 13 30			HARTFORD	C 7 LITCHFIELD 7.5 NE02	
28	09100	41 41 44	73 13 45			HARTFORD	C 7 LITCHFIELD 7.5 NE02	
29	09100	41 37 37	72 58 41			HARTFORD	C 9 BRISTOL 7.5 NE04 NE01 NE02	
30	09100	41 42 1	72 10 57	02	G 63526	HARTFORD	C15 WILLIMANTIC 7.5 NE05	
31	09101	41 33 1	73 24 21			HARTFORD	D 5 NEW MILFORD 7.5 NE02	
32	09101	41 32 38	73 24 23			HARTFORD	D 5 NEW MILFORD 7.5 NE02	
33	09101	41 33 55	73 24 40			HARTFORD	D 5 NEW MILFORD 7.5 NE02	
34	09101	41 35 35	73 26 59			HARTFORD	D 5 NEW MILFORD 7.5 NE02	
35	09101	41 31 13	73 25 10			HARTFORD	D 5 NEW MILFORD 7.5 NE02	
36	09103	41 34 11	72 53 23			HARTFORD	D 9 COLCHESTER 7.5 NE04	
37	09103	41 32 44	72 52 47			HARTFORD	D 9 COLCHESTER 7.5 NE04	
38	09103	41 31 16	72 49 38			HARTFORD	D10 MERIDEN 7.5 NE04	
39	09103	41 31 45	72 51 46			HARTFORD	D10 MERIDEN 7.5 NE04	
40	09103	41 31 22	72 50 20			HARTFORD	D10 MERIDEN 7.5 NE04	
41	09103	41 31 5	72 49 35			HARTFORD	D10 MERIDEN 7.5 NE04	
42	09107	41 32 30	72 33 13	02	K 63531	HARTFORD	D12 MIDDLE HADDAM 7.5 NE04	
43	09107	41 33 11	72 26 57	02	K 63532	HARTFORD	D13 MOODUS 7.5 NE04	
44	09111	41 33 31	72 7 19	02	G 63527	HARTFORD	D16	
45	09101	41 28 55	73 24 32			HARTFORD	E 5 DANBURY 7.5 NE02	
46	09101	41 27 22	73 23 44			HARTFORD	E 5 DANBURY 7.5 NE02	
47	09101	41 24 43	73 24 51			HARTFORD	E 5 DANBURY 7.5 NE02	
48	09101	41 24 24	73 25 32			HARTFORD	E 5 DANBURY 7.5 NE02	
49	09101	41 24 1	73 25 44			HARTFORD	E 5 DANBURY 7.5 NE02	
50	09101	41 23 12	73 25 26			HARTFORD	E 5 DANBURY 7.5 NE02	

Figure 10

The plot on this page was prepared by a digital plotter. Only DO data collected in the summer for the Hudson R. is plotted. The two extreme lines with symbols show 15th and 85th percentiles. The middle line shows the median or 50th percentile value. For additional examples of this program, see Section IV.

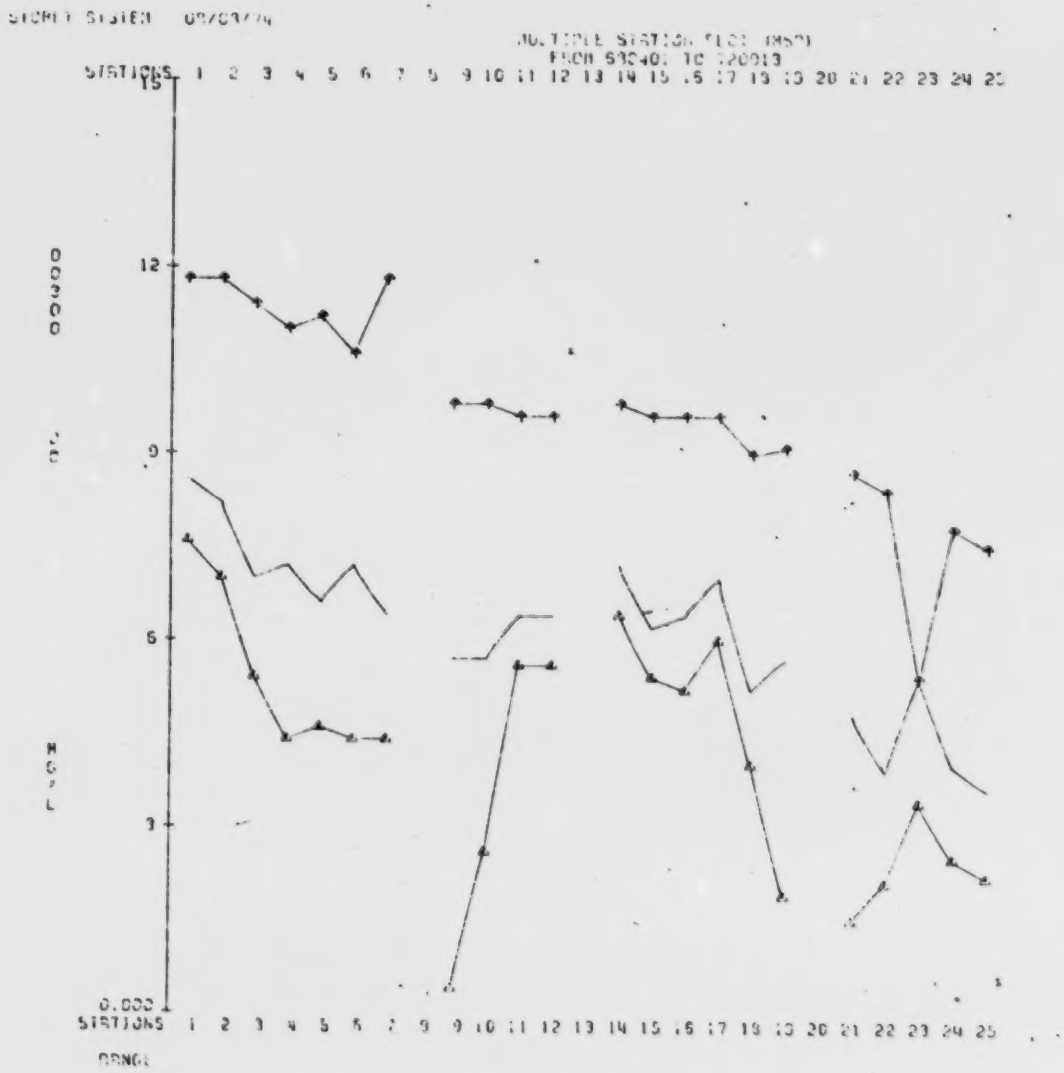


Figure 11

Figure 2. Sample Outputs From Multiple Station Plot Program. The first plot, scaled in miles on the horizontal axis, shows DO values in the Saginaw R. from its mouth to confluence with the Cass River and then upstream on the Cass R. for several miles. All symbols shown on the plot are medians (50th percentiles); lines were added manually for a quicker visual understanding of the data. Plots on the next pages show seasonal grouping of DO data for the Hudson R.

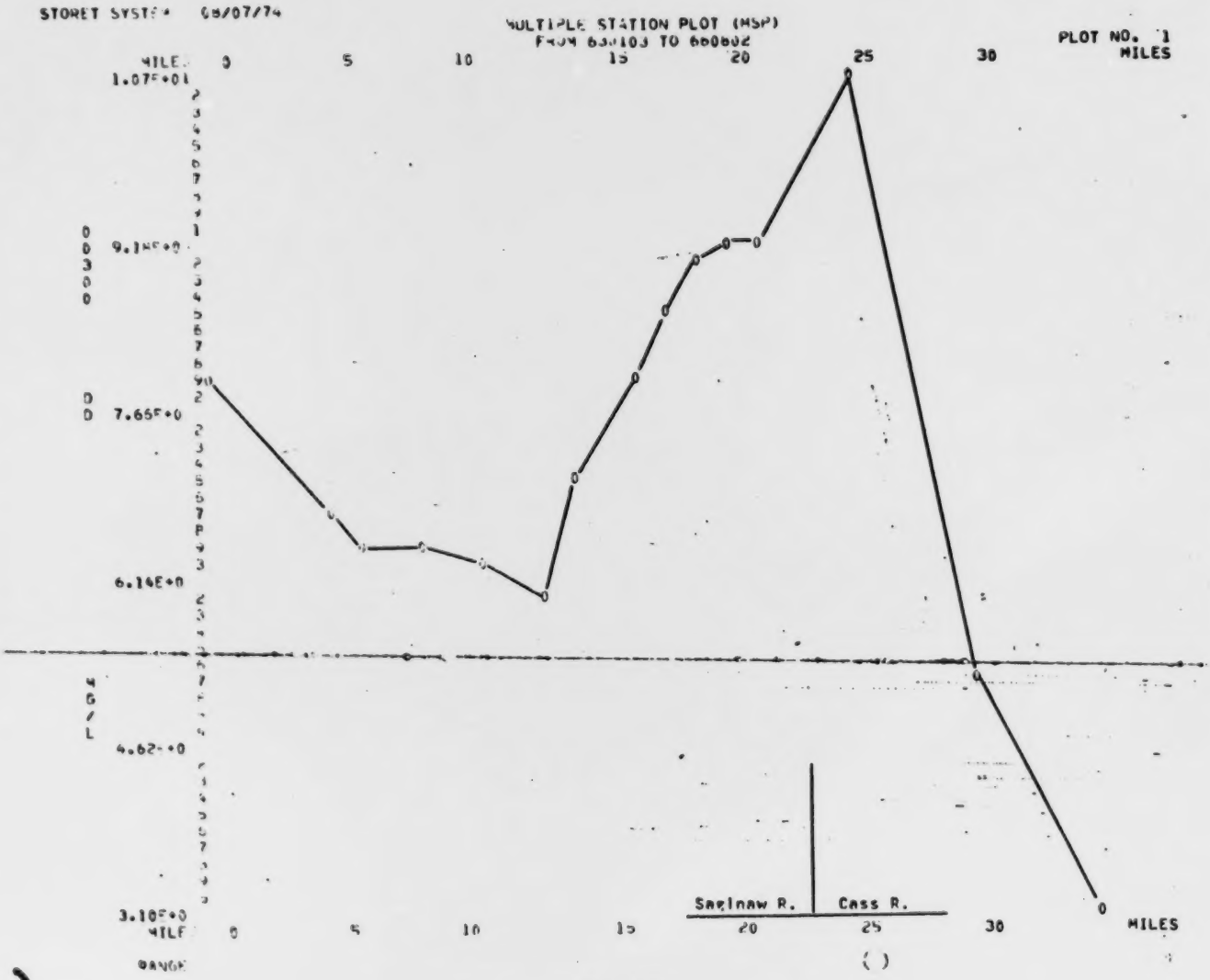


Figure 12

Plot on this page shows seasonal grouping of DO data taken since 1968. Stations numbered sequentially on the horizontal axis reference agency and station numbers shown on another page of output. Stations are in the correct river mile order, and, in this example, downstream (towards New York City) is to the right.

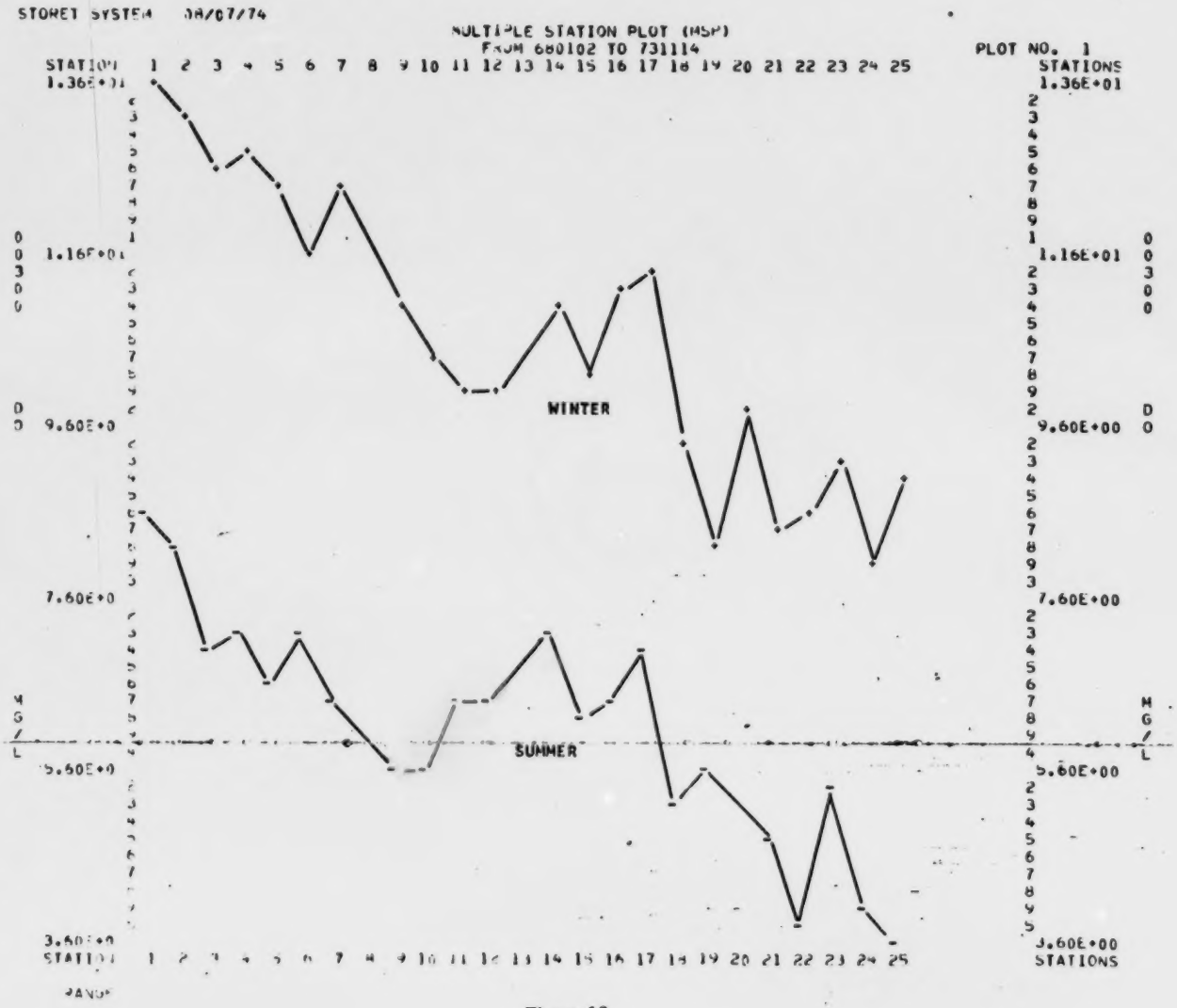


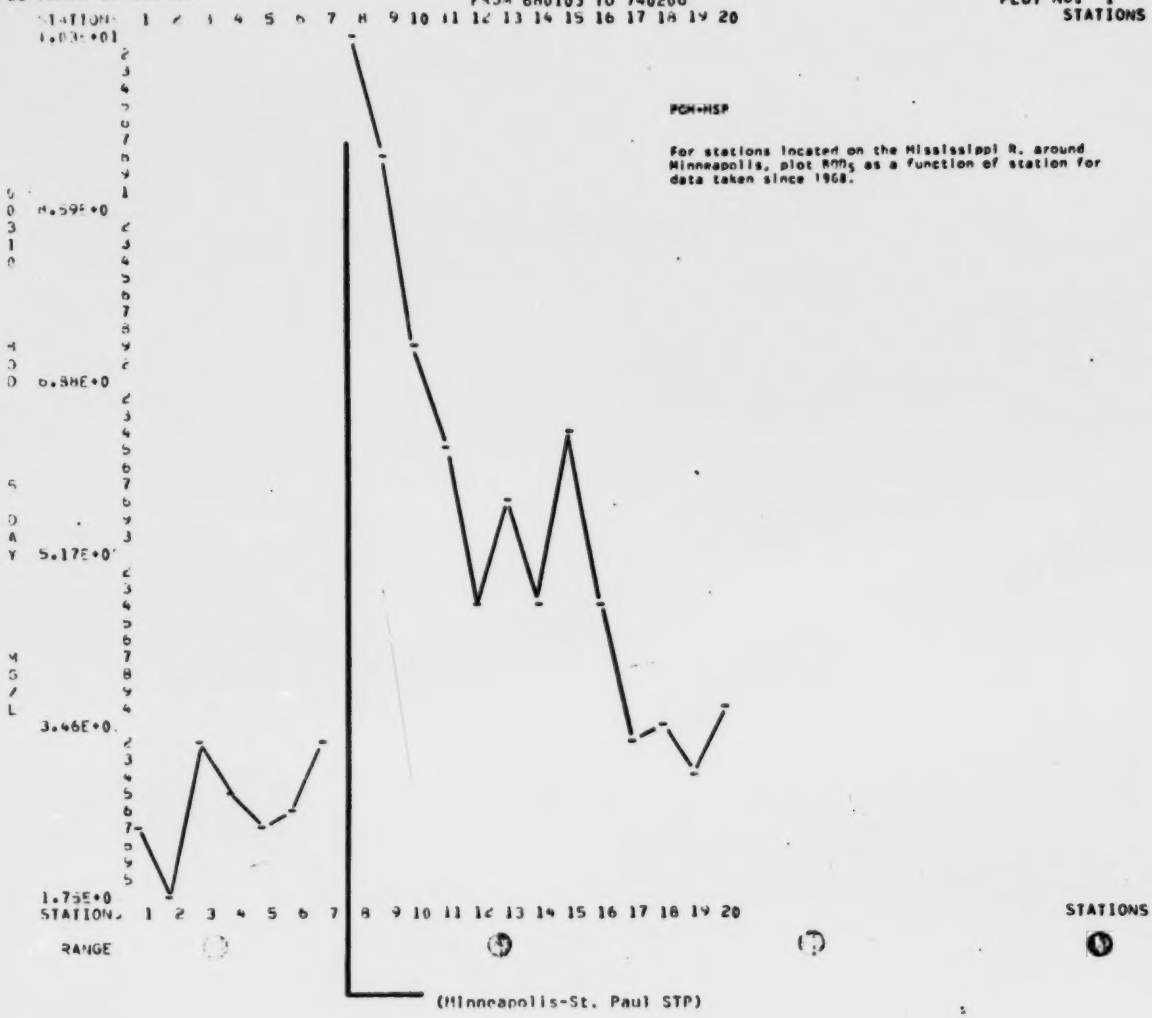
Figure 13

STORET SYSTEM 08/05/74

SAMPLE PAGES OF OUTPUT:

MULTIPLE STATION PLOT (MSP)
FROM 680103 TO 740206

PLOT NO. 1
STATIONS



PCM-MSP

For stations located on the Mississippi R. around Minneapolis, plot BOD_5 as a function of station for data taken since 1968.

STORET SYSTEM 08/05/74

MULTIPLE STATION PLOT (MSP)
FROM 680103 TO 740206

PLOT NO. 1

STATION	USER-STATION NO.	NP1	FROM	TO	NP2	FROM	TO
1	21MINN MSU-914C[illegible]15E67	61	680208	740115			
2	112[illegible] 052[illegible]	2	721 04	721124			
3	21MINN MSU-[illegible]---01E53	39	680228	730803			
4	1110 50 269043	36	680118	701022			
5	1110 50 269041	38	680118	701215			
6	1110 50 269039	115	680103	701015			
7	1110 50 269037	13	680829	701002			
8	1110 50 269035	13	680829	701002			
9	1110 50 269032	14	680829	701002			
10	1110 50 269030	95	680111	701030			
11	1110 50 269028	89	680111	701030			
12	1110 50 269026	14	680829	701002			
13	1110 50 269024	14	680829	701002			
14	1110 50 269022	95	680111	701030			
15	21MINN MSU-815-[illegible]15E58	29	680227	740[illegible]			
16	1110 50 269020	89	680111	701221			
17	1110 50 269013	85	680111	701030			
18	1110 [illegible] 260 27	70	680110	[illegible]			
19	112[illegible] [illegible]	81	690[illegible]13	731219			
20	21MINN MSU-71[illegible]-[illegible]15E62	63	[illegible]	740157			

Figure 14

SAMPLE PAGE OF OUTPUT:

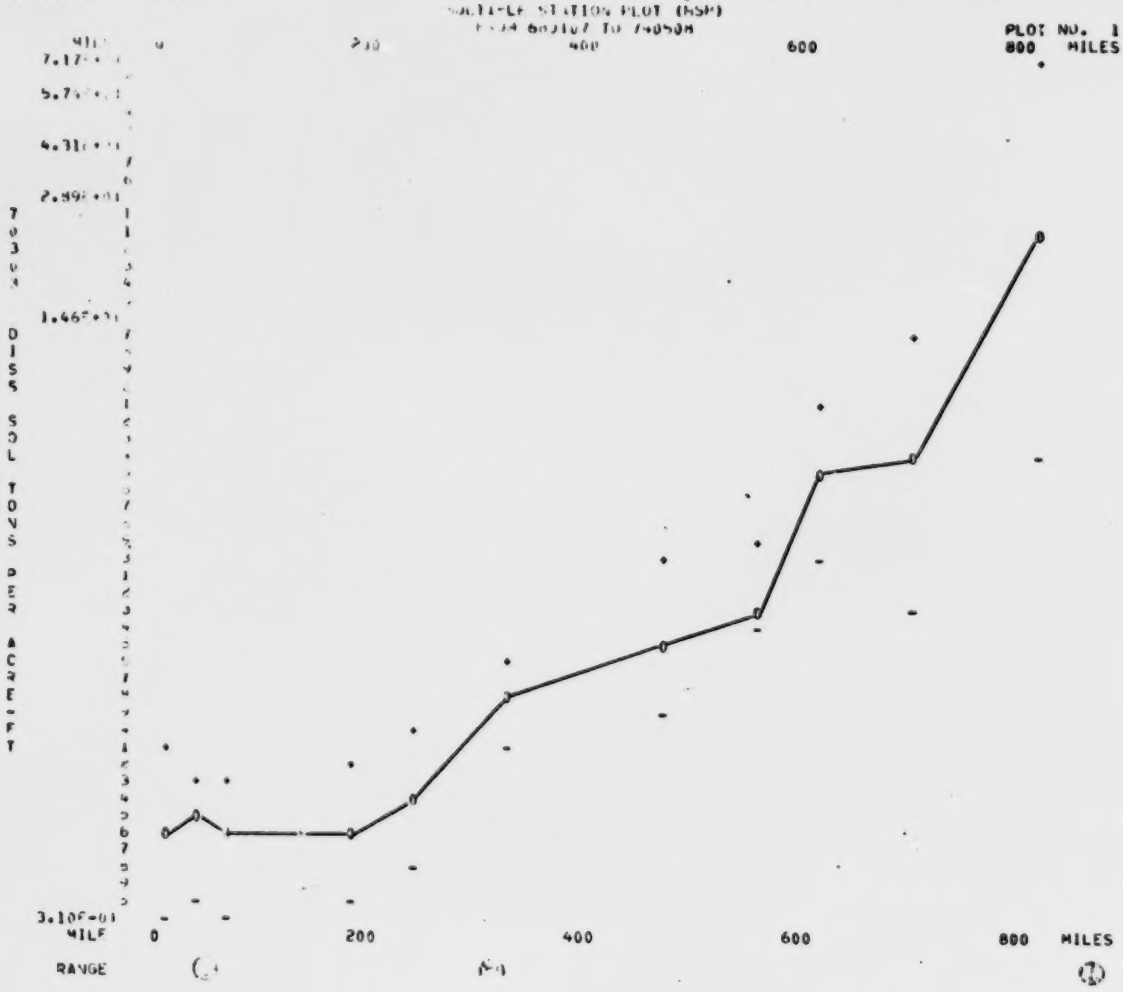
PGM=MSP

For stations located on the Brazos R., plot dissolved solids vs. stations. Use log scale for dissolved solids and linear mileage scaling for stations. Group upper stations on Salt Fk. of Brazos and lower reservoir stations. Begin analysis in 1968 and plot data only if the station has more than 30 observations in this time period. Also make a digital plot of the same data.

STORET SYSTEM 08/07/74

MULTIPLE STATION PLOT (MSP)
FROM 680107 TO 740508

PLOT NO. 1



STORET SYSTEM 08/07/74

MULTIPLE STATION PLOT (MSP)
FROM 680107 TO 740508

PLOT NO. 1

MILE	USGS-STATION NO.	NP1	FROM	TO	NP2	FROM	TO	STATION LOCATION
18.00	112[illegible] 0811720[illegible]	104	680120	740118				BRAZOS RIVER AT BRAZORIA RES
48.00	112[illegible] 0811665	147	681[illegible]3	740402				BRAZOS R NR ROSHARON, TEX.
74.00	112[illegible] 0811[illegible]	133	680111	740215				BRAZOS R AT RICHMOND, TEX.
204	112[illegible] [illegible]	102	680107	740226				BRAZOS R NR COLLEGE STATION,
264	112[illegible] [illegible]	129	680118	740415				BRAZOS R NR HIGHBANK, TEX.
304	112[illegible] [illegible]	* 20	70[illegible]	740504				BRAZOS R AT WACO, TEX.
348	112[illegible] [illegible]	49	680131	740227				BRAZOS R AT WHITNEY DAM NR W
478	112[illegible] [illegible]	41	701[illegible]	740213				BRAZOS RIVER NEAR DENNIS, TE
568	112[illegible] [illegible]	45	680131	740224				BRAZOS R AT POSSUM KINGDOM D
627	112[illegible] [illegible]	33	691[illegible]	740225				BRAZOS R AT FM 1287 NR GRAHA
643	112[illegible] [illegible]	* 1	72[illegible]	72[illegible]				BRAZOS R NR SOUTH BEND, TEX.
6[illegible]	112[illegible] [illegible]	37	691[illegible]	740225				BRAZOS RIVER NEAR GRAHAM, TE
7[illegible]	112[illegible] [illegible]	171	680117	731[illegible]				BRAZOS RIVER AT SEYMOUR, TEX
816	112[illegible] [illegible]	244	680118	740117				SALT FK BRAZOS R NR ASPERMON

Figure 15

Reproduced from best available copy.

SAMPLE PAGE OF OUTPUT: PROM-MSP

For selected stations in Ft. Loudoun Reservoir, Tenn., and upstream and downstream points, plot DO since 1970. Group DO measurements taken above 10.1 feet and measurements below 10.1 feet. Use linear mileage scaling.

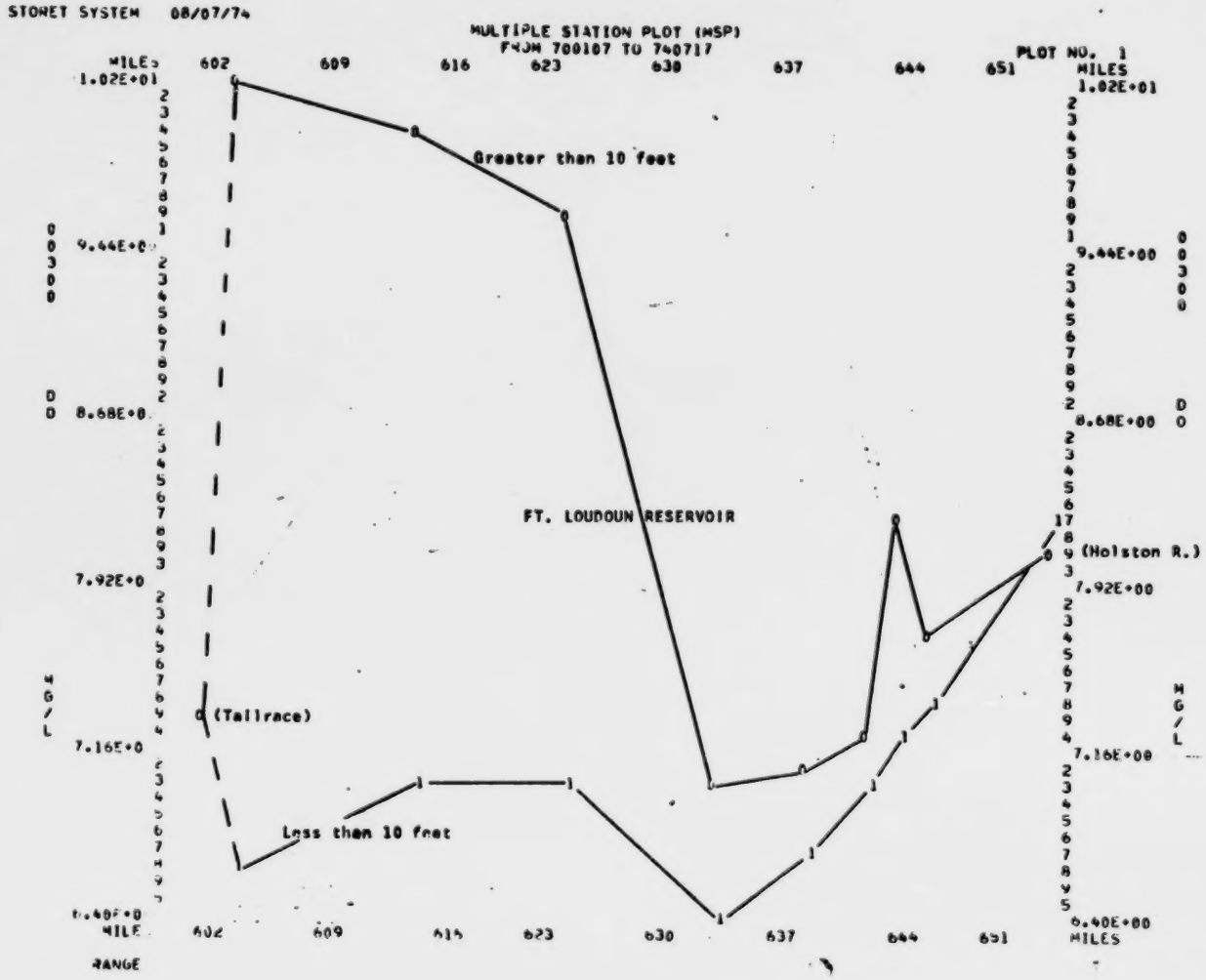
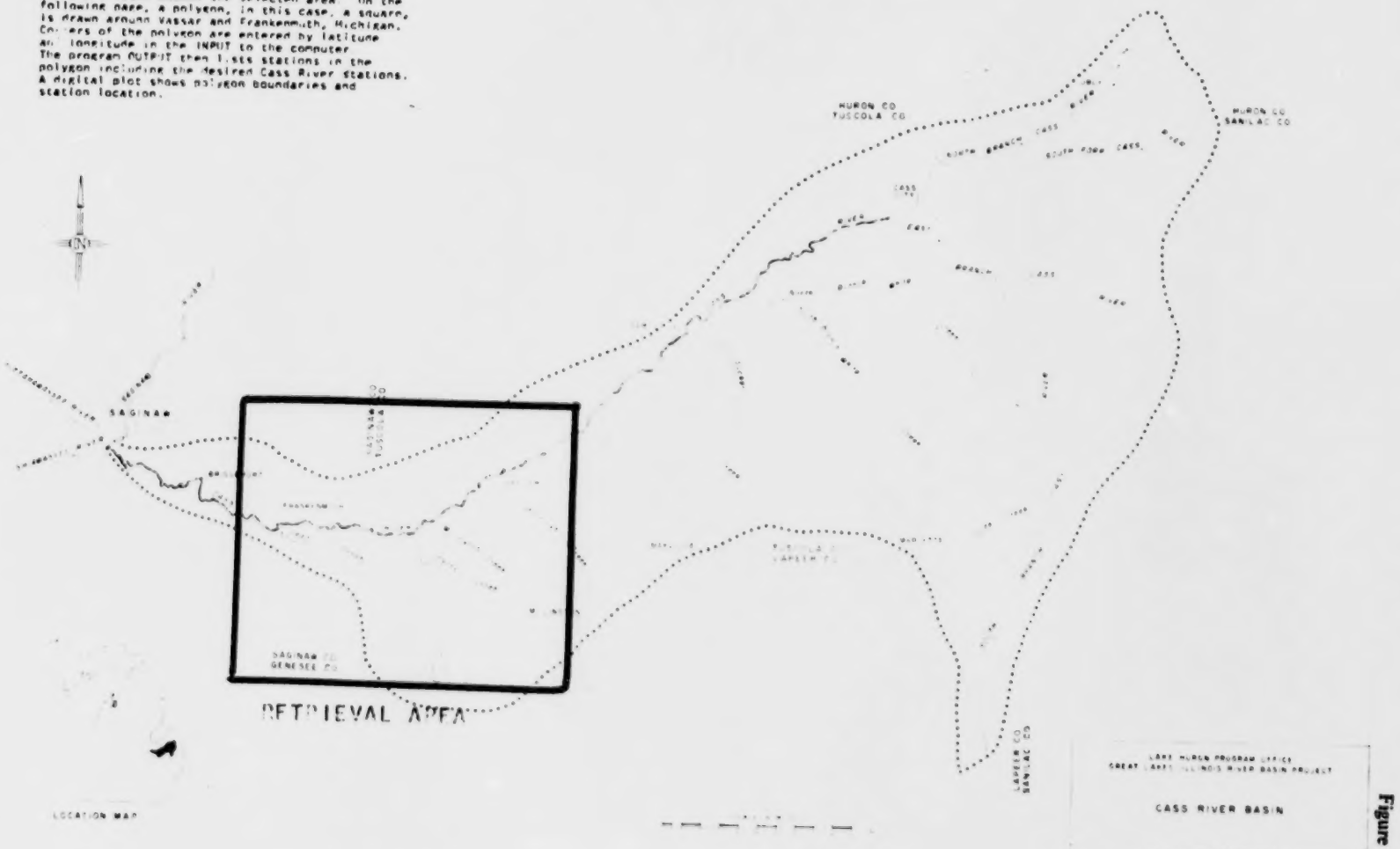


Figure 16

PGM=LOC

If an analyst wishes to retrieve stations having lat/long that lie within a selected area, he first draws a polygon around the selected area. On the following page, a polygon, in this case, a square, is drawn around Vassar and Frankenmuth, Michigan. Corners of the polygon are entered by latitude and longitude in the INPUT to the computer. The program OUTPUT then lists stations in the polygon including the desired Cass River stations. A digital plot shows polygon boundaries and station location.



LAKE HURON PROGRAM OFFICE
GREAT LAKES ILLINOIS RIVER BASIN PROJECT

CASS RIVER BASIN

U. S. DEPARTMENT OF THE INTERIOR
FEDERAL WATER POLLUTION CONTROL ADMINISTRATION
GREAT LAKES BASIN - UNIVERSITY OF MICHIGAN

Figure 17

INPUT:

```
MOM RET.LOC.MM: 60.1405.SCHT 2000.00.
./PLIR4    30: 19.12.43  9:[illegible]  '2.01
LT 1.L .3124.L  3500.L 43..L 0310.L 4320.L 0100.L 4   .L 0350.
NM 13.
```

OUTPUT:

```
AO    2134114 .00332        43 14 00.0 083 33 00.0    0 PRODUCING WELL                 26    MICHIGAN
BO    2134114 .00333        43 14 00.0 083 34 00.0    0 PRODUCING WELL                 26    MICHIGAN
CO    2134114 .00124        43 16 00.0 083 43 00.0    0 PRODUCING WELL                 26    MICHIGAN
DO    2141CM  73187         43 19 17.0 083 48 00.0    0 CASS R. 3.5 MILES W. OF FRANKENM 26  SAGINAW CO.
EO    1115305 .03452        43 19 27.0 083 34 25.0    0 CASS RIV. DASAU-L.M.P.O.  A640   26    MICHIGAN
FO    1115305 .03482        43 19 27.0 083 34 25.0    0 CASS DASAU--PLANT CR.  LMPO A642 26    MICHIGAN
GO    1115305 .03431        43 19 27.0 083 44 25.0    0 CASS RIV. DASAU-L.M.P.O.  A630   26    MICHIGAN
HO    2141CM  73094         43 19 28.0 083 44 27.0    0 CASS RIVER AT M-81.            26    SAGINAW CO.
IO    1115305 .03713        43 19 30.0 083 45 00.0    0 FRANKENMUTH STP      LMPO A628 26    MICHIGAN
JO    1115305 .03456        43 19 40.0 083 46 32.0    0 CASS RIV. DASAU-L.M.P.O.  A620   26    MICHIGAN
KO    2141CM  73099         43 19 41.0 083 45 27.0    0 CASS RIVER AT DEHMEL ROAD.     26    SAGINAW CO.
LO    2141CM  73100         43 19 44.0 083 46 38.0    0 CASS RIVER AT S. HELVER RD.    26    SAGINAW CO.
MO    1115305 .03444        43 19 43.0 083 49 30.0    0 CASS RIV. DASAU-L.M.P.O.  A615   26    MICHIGAN
NO    2141CM  73071         43 19 48.0 083 49 49.0    0 CASS R AT DIXIE HWY BRIDGE     26    SAGINAW COUNTY
OO    112440  .4131500      43 19 43.0 083 46 33.0    0 CASS RIVER AT FRANKENMUTH, MICHI 26
PO    1115305 .03440        43 21 00.0 083 35 00.0    0 CASS DASAU-GOODINGS CR.LMPO A649 26  MICHIGAN
QO    1115305 .03714        43 21 53.0 083 35 12.0    0 VASSAR STP           LMPO A647 26    MICHIGAN
RO    1115305 .03453        43 22 10.0 083 34 35.0    0 CASS RIV. DASAU-L.M.P.O.  A630   26    MICHIGAN
SO    112440  .4151000      43 22 10.0 083 34 35.0    0 CASS R. AT VASSAR, MICH.       26
TO    2141CM  74043         43 22 13.0 083 34 22.0    0 CASS RIVER AT M-15.            26    TUSCOLA CO.
UO    1115305 .03454        43 23 32.0 083 36 30.0    0 CASS RIV. DASAU-L.M.P.O.  A655   26    MICHIGAN
VO    2141CM  74042         43 23 34.0 083 38 33.0    0 CASS RIVER AT WATERMAN ROAD.   26    TUSCOLA CO.
WO    2134114 .00089        43 24 00.0 083 43 00.0    0 PRODUCING WELL                 26    MICHIGAN
LAT. MAX.443 23 17.9
LAT. MIN.443 1  42.0
LONG. MAX.* 83 31 59.9
LONG. MIN.* 83 28  0.0
```

(double letters referenced to tags on digital plot, next page)

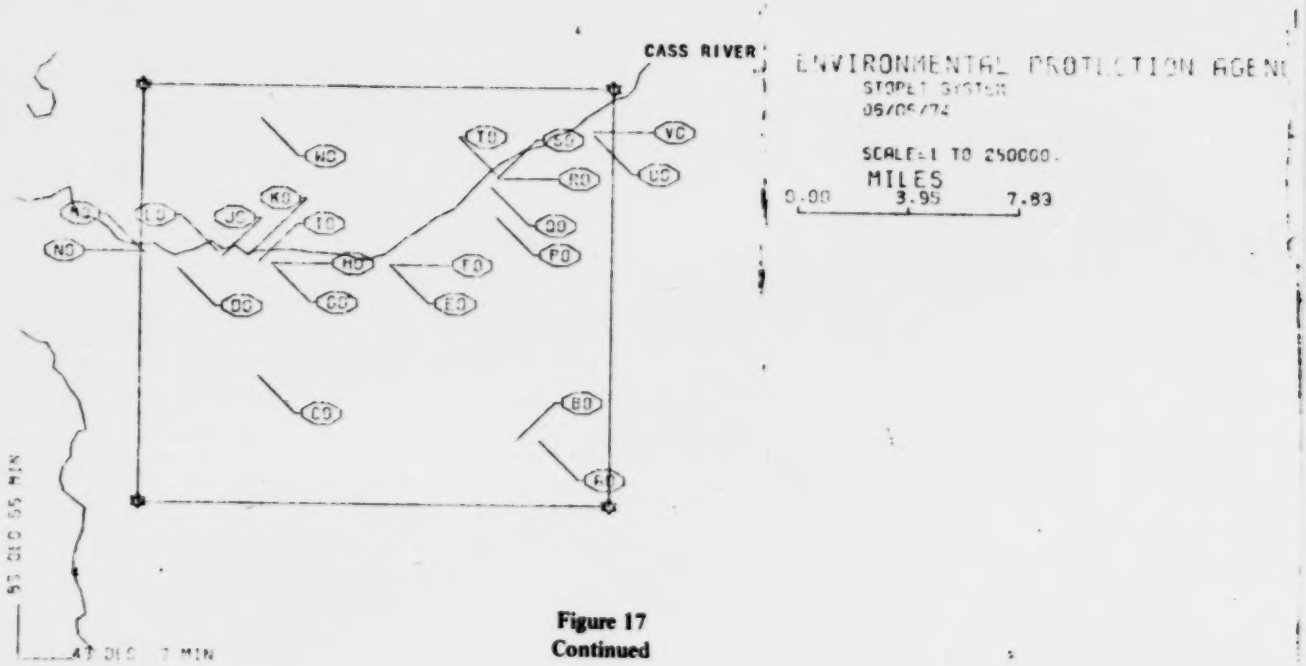
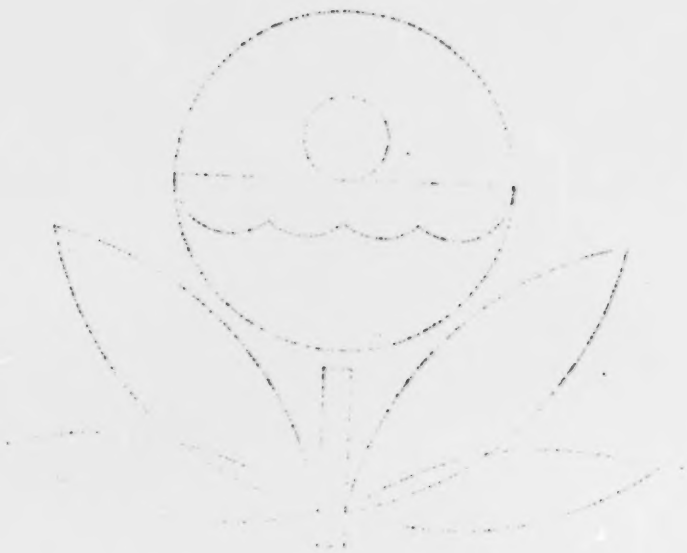


Figure 17
Continued



CALL FLOWER (X,Y,R,A)

Load module on
BCS011,TOBIN,LOAD.SUBLIB

X is x - coordinate, in inches,of center of large circle
Y is y - coordinate, " " , " " " " "
R is radius of large circle, in inches (1" used above)
A is angular resolution, in radius (.05 used above)

all floating point
single precision

Figure 18

Figure 19
Base Map of Continental United States

Figure 20
Outline Map of Florida

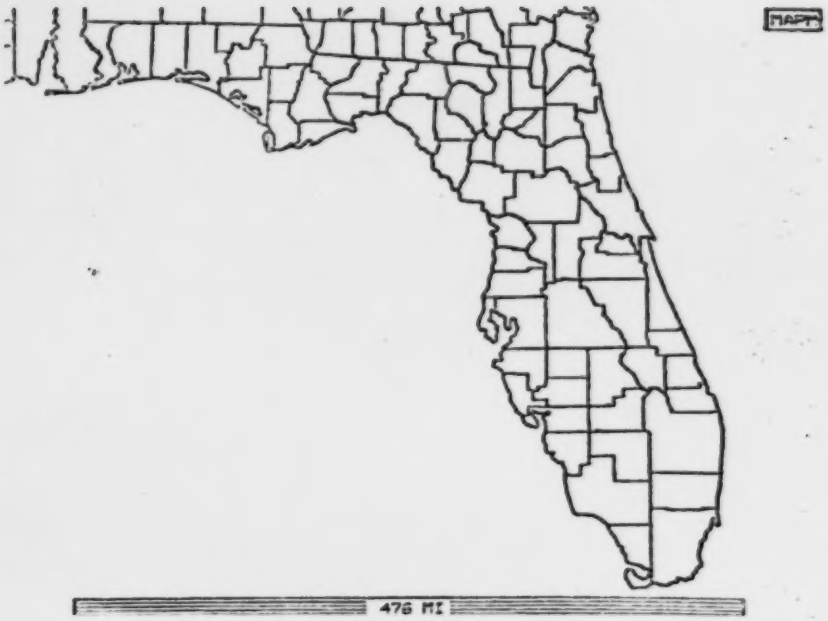


Figure 21
County Overlay for Florida

Figure 22
Hydrological Features for Florida

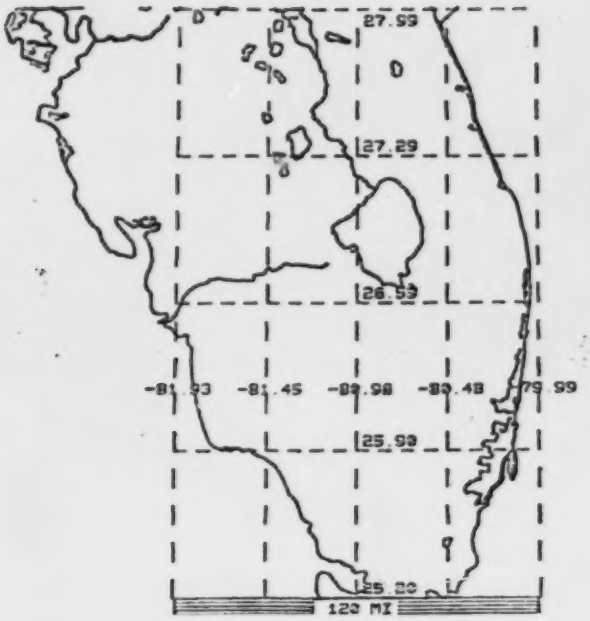


Figure 23
Zoom-Derived Map of Southern Florida with Geographic Grid

Figure 24
Water Quality Stations in Michigan

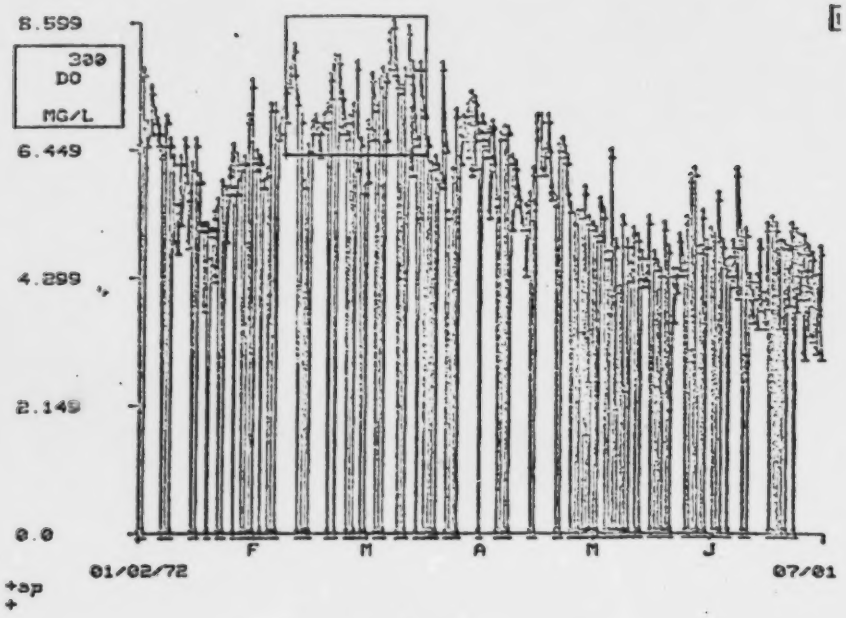


Figure 25
Graphical Retrieval of Dissolved Oxygen Data

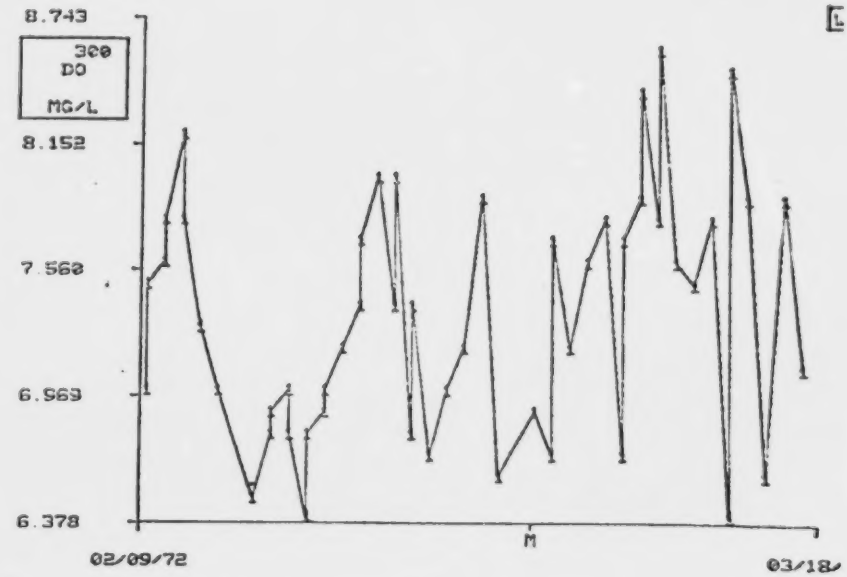


Figure 26
Data-Zoom Graph for Dissolved Oxygen

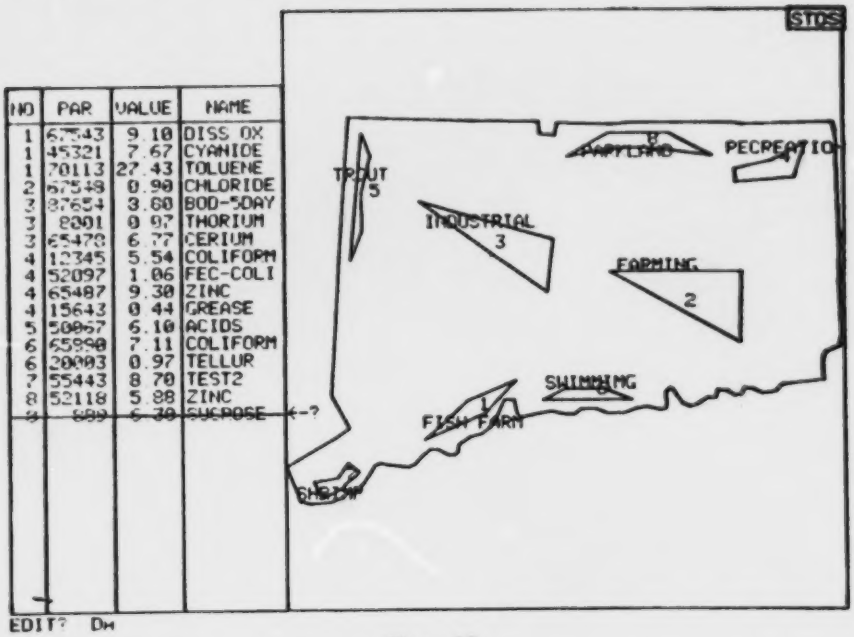


Figure 27
Example of Standards Zones Display

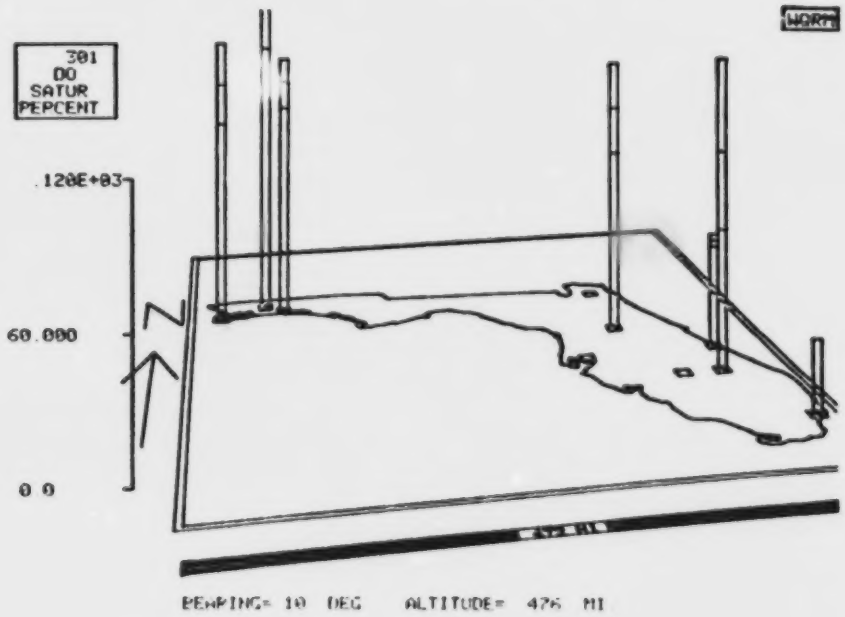


Figure 28
Example of Perspective Data Display

Figure 29
Graphical Output from AUTOMAT Subsystem

Figure 30
Graphcial Output from Municipal Waste Inventory

A METHOD FOR IMPROVING USER ACCESS TO STORET

By Michael J. Friedland

INTRODUCTION

Several projects at the National Environmental Research Center in Las Vegas, Nevada (NERC-LV), require that collected data be stored in the STORET information system. There is also a need to perform statistical analysis of the data using packages such as SAS, BIOMED, etc. User personnel have limited data processing experience (a maximum of one semester of FORTRAN), and no experience (or interest) in IBM job control language (JCL). However, they are willing to learn a minimum of STORET and WYLBUR commands.

In order to make the STORET-resident data available for manipulation, a FORTRAN program has been written to read a MORE file and write the results back out in card-image format; a subsequent step puts the file into EDIT format so the WYLBUR user can access and save it easily. The user then builds the JCL required to run SAS around the data set and submits it via the RUN command. The details of this procedure are illustrated in the figures.

Figure 1

The data set RUN.LAKE.READ contains the JCL and data needed to perform a STORET retrieval, run the program to convert the MORE file from packed EBCDIC to card-image form, and then convert the output to EDIT format. The user is instructed to USE the data set, alter it as necessary (change STORET input or output data set name) and before RUNning it, SAVE it under the name of the output data set (in this case, RET.DATA) to validate the DISP=OLD clause on the EXEC EDSCARDS statement. The only other item the user must attend to is the control input to the reformatting program; that is, the first two columns of the one input card must contain the number of parameters retrieved.

Figure 2

The reformatting program produces the data set shown in this figure. The data set has the following attributes:

- Card image
- Six values per card
- Null values are indicated by −123.000
- Identification information (station-date-time-depth) appears in columns 61-79. A one digit sequence number is in column 80.

The SAS user typically performs a

CHANGE'-123.000' TO' 'IN ALL NOLIST

command to convert the null value indicator to a blank field, as is most commonly required by SAS.

Figure 3

The mainline program used in this procedure is shown here. It initializes subroutine GETDAT, the core of this procedure, and performs the output operation.

Figure 4

A listing of subroutine GETDAT is presented here. This subroutine has proved to be generalized enough to be used by several different applications of reading STORET MORE files. The commentary at the top of the routine is intended to communicate all necessary information to the user writing a mainline to call this subprogram.

```
 1.      //RNSWY JOB (A805,HIN,01,30,30),' STR-TO-WYL',MSGLEVEL=(0,0)
 2.      // EXEC WQDIST
 3.      //DIST.CARDFD DD *
 4.      PGM RET,A 11EPALES,MORE=4,RMT 35,B P,
 5.      P 610,DR 3,P 32217,P 10,P 77,
 6.      P 665,DR 4,P 671,P 630,DR 3,P 300,P 72025,
 7.      P=8,P=3,P=85508,
 8.      S=270000,S=279999,
 9.      S=550000,S=559999,
10.      BD=720601,ED=720731,
11.      PRT=NO,
12.      HEAD=RET111,
13.      // EXEC EDSIN,NAME='CNA805.RNS.LAKE.READ.SOURCE',DISK=TS0004,
13.5     // INPUT='&INP1'
14.      // EXEC EDSIN,NAME='CNLV03.MJF.GETDAT.SOURCE',DISK=TS0004,
14.5     // INPUT='&INP2'
15.      // EXEC FORTGCLG,PARM='NOSOURCE,NOMAP'
15.5     //FORT.SYSIN DD DSN=&INP1,DISP=SHR
16.      //  DD DSN=&INP2,DISP=SHR
18.      //GO.FT15F001 DD DSN=&FCF,DISP=(OLD,DELETE)
19.      //GO.FT12F001 DD DSN=&LAKFIL,DISP=(NEW,PASS),UNIT=SYSDA,
20.      //  SPACE=(TRK,(25,10)),DCB=(RECFM=FB,LRECL=80,BLKSIZE=1680)
21.      //GO.SYSIN DD *
22.      12  NUMBER OF PARAMETERS RETRIEVED IN 1-2 (1-50)
23.      // EXEC EDSCARDS,NAME='CNA805.RNS.RET.DATA',DISK=TS0004,TYPE=OLD
24.      //CARDS DD DSN=&LAKFIL,DISP=OLD
```

Figure 1

						STA	DATE TIME	DEPTH SEQ
1.	0.050	-123.000	23.900	-123.000	0.010	-123.00027150372	7111930	41
2.	0.040	8.800	-123.000	1203.000	4.000	-123.00027150372	7111930	42
3.	0.050	-123.000	20.900	-123.000	0.011	-123.00027150372	7111930	151
4.	0.040	9.200	-123.000	1202.000	15.000	-123.00027150372	7111930	152
5.	0.050	-123.000	19.400	-123.000	0.010	-123.00027150372	7111930	211
6.	0.040	8.400	-123.000	1201.000	21.000	-123.00027150372	7111930	212
7.	0.080	9.000	21.600	38.000	0.024	-123.00027160172	7 316 5	01
8.	0.060	9.800	-123.000	1089.000	0.0	-123.00027160172	7 316 5	02
9.	0.070	-123.000	20.400	-123.000	0.032	-123.00027160172	7 316 5	101
10.	0.060	7.600	-123.000	1090.000	10.000	-123.00027160172	7 316 5	102
11.	0.090	-123.000	17.100	-123.000	0.033	-123.00027160172	7 316 5	151
12.	0.050	1.400	-123.000	1091.000	15.000	-123.00027160172	7 316 5	152
13.	0.150	-123.000	14.000	-123.000	0.026	-123.00027160172	7 316 5	221
14.	0.080	0.200	-123.000	1092.000	22.000	-123.00027160172	7 316 5	222
15.	0.360	-123.000	11.400	-123.000	0.032	-123.00027160172	7 316 5	301
16.	0.070	1.000	-123.000	1093.000	30.000	-123.00027160172	7 316 5	302
17.	0.060	6.200	22.600	36.000	0.043	-123.00027160272	7 317 5	01
18.	0.040	10.000	-123.000	1094.000	0.0	-123.00027160272	7 317 5	02
19.	0.080	-123.000	22.600	-123.000	0.025	-123.00027160272	7 317 5	151
20.	0.040	0.200	-123.000	1096.000	15.000	-123.00027160272	7 317 5	152
21.	0.750	-123.000	12.300	-123.000	0.032	-123.00027160272	7 317 5	221
22.	0.040	5.800	-123.000	1095.000	22.000	-123.00027160272	7 317 5	222
23.	0.100	-123.000	15.300	-123.000	0.025	-123.00027160272	7 317 5	261
24.	0.050	0.600	-123.000	1097.000	26.000	-123.00027160272	7 317 5	262
25.	0.410	-123.000	10.500	-123.000	0.059	-123.00027160272	7 317 5	281
26.	0.070	0.600	-123.000	1098.000	28.000	-123.00027160272	7 317 5	282
27.	0.090	-123.000	21.900	36.000	0.029	-123.00027160372	7 51810	01
28.	0.080	8.200	-123.000	1099.000	0.0	-123.00027160372	7 51810	02
29.	0.040	2.600	21.000	160.000	0.010	-123.00027170172	7 52045	01
30.	0.040	9.400	-123.000	1038.000	0.0	-123.00027170172	7 52045	02
31.	0.030	-123.000	20.900	-123.000	0.010	-123.00027170172	7 52045	151
32.	0.040	9.200	-123.000	1039.000	15.000	-123.00027170172	7 52045	152
33.	0.020	-123.000	15.500	-123.000	0.011	-123.00027170172	7 52045	361
34.	0.030	6.400	-123.000	1040.000	36.000	-123.00027170172	7 52045	362
35.	0.070	-123.000	13.400	-123.000	0.016	-123.00027170172	7 52045	401
36.	0.050	4.000	-123.000	1041.000	40.000	-123.00027170172	7 52045	402
37.	0.140	-123.000	11.900	-123.000	0.017	-123.00027170172	7 52045	451
38.	0.060	1.400	-123.000	1042.000	45.000	-123.00027170172	7 52045	452
39.	0.020	2.800	20.800	160.000	0.012	-123.00027170272	7 52130	01
40.	0.040	9.200	-123.000	764.000	0.0	-123.00027170272	7 52130	02
41.	0.020	-123.000	20.400	-123.000	0.010	-123.00027170272	7 52130	151
42.	0.030	9.200	-123.000	775.000	15.000	-123.00027170272	7 52130	152
43.	0.010	-123.000	15.800	-123.000	0.010	-123.00027170272	7 52130	331
44.	0.030	5.200	-123.000	776.000	33.000	-123.00027170272	7 52130	332
45.	0.050	17.100	24.000	30.000	0.038	-123.00027190172	6301945	01
46.	0.110	16.800	-123.000	816.000	0.0	-123.00027190172	6301945	02
47.	0.170	-123.000	17.000	-123.000	0.053	-123.00027190172	6301945	151
48.	0.720	7.800	-123.000	815.000	15.000	-123.00027190172	6301945	152
49.	0.850	-123.000	10.000	-123.000	0.432	-123.00027190172	6301945	301
50.	0.250	0.400	-123.000	814.000	30.000	-123.00027190172	6301945	302
51.	0.050	32.800	25.000	28.000	0.042	-123.00027190272	6302015	01
52.	0.060	16.800	-123.000	819.000	0.0	-123.00027190272	6302015	02
53.	0.220	-123.000	17.000	-123.000	0.065	-123.00027190272	6302015	151
54.	0.900	6.200	-123.000	818.000	15.000	-123.00027190272	6302015	152
55.	0.680	-123.000	11.200	-123.000	0.241	-123.00027190272	6302015	271
56.	0.710	0.300	-123.000	817.000	27.000	-123.00027190272	6302015	272
57.	0.120	45.100	22.000	14.000	0.168	-123.00027200172	7 21030	01
58.	0.090	6.200	-123.000	1023.000	0.0	-123.00027200172	7 21030	02
59.	0.130	25.600	22.000	-123.000	0.054	-123.00027210172	7 31820	01
60.	0.130	12.000	-123.000	1086.000	0.0	-123.00027210172	7 31820	02

Figure 2

```
C   MAINLINE PROGRAM TO READ STORET MORE FILE AND OUTPUT IT IN
C   CARD IMAGE FORM.
      DIMENSION KARD(21)
      COMMON/GVALUE/VALUE(10)
      COMMON/GPARM /NDFIL,NUMRET,MORE,INFIL,IDNUL,VALNUL
      COMMON /GCODE/NSTA(16),NDATE(3)
     X   ,NTIME(2)
      COMMON /GDEEP/DEPTH
      LAKFIL=12
C  MORE=4 TO GET DEPTH
      MORE=4
C
C CALL SETNUM TO FIND NUMBER OF PARAMETERS RETRIEVED, AND IF
C  MORE = 3 RETRIEVAL WAS RUN.
C
      CALL SETNUM(NUMRET,MORE3)
C  CALCULATE NUMBER OF LINES PER SAMPLE
      LINES=1+NUMRET/6
      IF(MOD(NUMRET,6).EQ.0)LINES=LINES-1
      KASES=0
      VALNUL=-123.
      IF(MORE3.EQ.1)MORE=3
   50 CONTINUE
      CALL GETDAT
      IF(NDFIL.NE.0)GO TO 200
      KASES=KASES+1
C
C  WRITE 6 VALUES/LINE, F10.3  WITH STATION-DATE-TIME-DEPTH-SEQ IN 61-80
C
      IDEEP=DEPTH
      J1=1
      DO 90 I=1,LINES
      J2=J1+5
      WRITE(LAKFIL,85)(VALUE(J),J=J1,J2),NSTA(1),NSTA(2),NDATE,NTIME,
     X   IDEEP,I
      J1=J1+6
   85 FORMAT(6F10.3,A4,A2,5I2,I3,I1)
   90 CONTINUE
      GO TO 50
  200 ENDFILE LAKFIL
      REWINDLAKFIL
      STOP
      END
      SUBROUTINE  SETNUM(NUMRET,MORE3)
C
      DATA M3/'MOR3'/
C IF COLS 5-8 HAVE 'MOR3',SET MORE3=1 SO CALLING PROG SETS MORE=3
C
C READ A CARD IN I2 FORMAT, NUMRET IN 1,50.IF EOF,
C     DO NOT SET. IF INVALID, WRITE DIAGNOSTIC ON UNIT 12 AND STOP 154
C
      MORE3=0
      READ(5,1,END=100,ERR=9  )NUM,MORE3
    1 FORMAT(I2,2X,A4)
      IF(NUM.GE.1.AND.NUM.LE.50)GO TO 20
C
C  BAD NUM
C
    9 WRITE(12,10)NUM
   10 FORMAT(' NUMBER OF PARAMETERS NOT IN (1,50), RUN STOPPED', I3)
      STOP 154
C
C
C
   20 NUMRET=NUM
      IF(MORE3.EQ.M3)MORE3=1
  100 RETURN
      END
```

Figure 3

```
C                                                                       GTDA  10
C*  SUBROUTINE IS CALLED TO EXTRACT DATA FROM STORET 'MORE' FILE        GTDA  15
C     AND RETURN DATA VALUES VIA LABELLED COMMON. VARIOUS STATUS        GTDA  20
C     PARAMETERS ARE AVAILABLE; FUNCTIONS ARE EXPLAINED BELOW.          GTDA  25
C                                                                       GTDA  30
C*  THE COMMON BLOCK NAMES, SIZES AND USES ARE:                         GTDA  35
C                                                                       GTDA  40
C   1.GVALUE -- 50 WORDS,FLT.PT.; DATA VALUES                           GTDA  45
C                                                                       GTDA  50
C   2.GNULL  -- 50 WORDS,INTEGER; NULL VALUE INDICATORS                 GTDA  55
C                                                                       GTDA  60
C   3.GCODE  -- 21 WORDS,INTEGER; STATION ID(16),DATE(3),TIME(2)        GTDA  65
C                                                                       GTDA  70
C   4.GKNTR  -- 52 WORDS,INTEGER; COUNTS OF STATIONS,RECORDS, AND NULLS GTDA  75
C                                                                       GTDA  80
C   5.GPARM  --  6 WORDS:5 INT,1 FL.PT.; EOF FLAG, OTHER PARAMS         GTDA  85
C                                                                       GTDA  90
C   6.GFORM  -- 11 WORDS,INTEGER; PROVIDES FOR ALTERNATE STATION ID FMT GTDA  95
C                                                                       GTDA 100
C   7.GECHO  --  1 WORD,FLT.PT.,,PRESET TO ZERO. IF SET TO 1,CAUSES     GTDA 105
C       GETDAT TO ECHO EACH LINE READ                                   GTDA 110
C   8.GDEEP  --  1 WORD FLT.PT., DEPTH. CHARS 300-304 OF MORE=4 DATA
C        RECORD.
C
C                                                                       GTDA 115
C   THE COMMON BLOCK VARIABLES'-FUNCTIONS AND VALUES FOLLOW             GTDA 120
C                                                                       GTDA 125
C                                                                       GTDA 130
C                                                                       GTDA 135
C                                                                       GTDA 140
C  VALUE(I) CONTAINS THE QUANTITY FOR THE I-TH PARAMETER RETRIEVED.     GTDA 145
C     IF THE VALUE IS NULL, VALUE(I) HAS 1.E-15. THE USER MAY           GTDA 150
CC    CHOOSE A DIFFERENT DEFAULT BY SETTIING VALNUL, THE 6TH (FL.PT.)   GTDA 155
C     WORD OF BLOCK GPARM, TO THE DESIRED QUALITY. THE ARRAY VALUE      GTDA 160
C     IS SET BY EACH CALL TO GETDAT; IT IS NOT PRESET.                  GTDA 165
C                                                                       GTDA 170
C  NULL(I) IS SET TO 10 IF VALUE(I) IS NULL. THE USER MAY CHOOSE A      GTDA 175
C     DIFFERENT INDICATOR VALUE BY SETTING  IDNUL, THE 5TH WORD(INTGR)  GTDA 180
C     OF BLOCK GPARM, TO THE DESIRED VALUE. THE ARRAY NULL IS SET       GTDA 185
C     BY EACH CALL TO GETDAT; IT IS NOT PRESET.                         GTDA 190
C                                                                       GTDA 195
C                                                                       GTDA 200
C  THE STATION ID IS RETURNED IN THE FIRST 3 WORDS OF THE ARRAY NSCODE  GTDA 205
C     IN THE FORMAT A4,A3,A2. THE USER MAY SUPPLY AN ALTERNATE FORMAT   GTDA 210
C     AS DESCRIBED UNDER BLOCK GCFORM.                                  GTDA 215
C  THE DATE IS READ AS 3I2 (YEAR,MONTH,DAY).THE TIME IS 2I2 (HOUR,MIN)  GTDA 220
C                                                                       GTDA 225
C                                                                       GTDA 230
C* THE VARIABLE KNTSTA IS INCREMENTED WHEN A NEW STATION IS ENCOUNTERED.GTDA 235
C     IT IS PRESET TO ZERO.                                             GTDA 240
C                                                                       GTDA 245
C* THE VARIABLE KNTREC IS INCREMENTED WHEN A LINE OF DATA,NOT END-OF-   GTDA 250
C     -FILE, IS READ. IT IS PRESET TO ZERO.                             GTDA 255
C                                                                       GTDA 260
C* THE VARIABLE KNTNUL(I) IS INCREMENTED WHEN VALUE(I) IS NULL.         GTDA 265
C     THE ARRAY IS PRESET TO ZERO.                                      GTDA 270
C                                                                       GTDA 275
C                                                                       GTDA 280
C* ---VARIABLE-------USAGE---------------------DEFAULT------SET BY----  GTDA 285
C      NOMORE       1 IF EOF                    ZERO         GETDAT     GTDA 290
```

Figure 4

```
C       NUMRET       NO. OF PARAMS IN RETRIEVAL   10            USER        GTDA 295
C       MORE         TYPE OF MORE FILE (3 OR 4)    3            USER        GTDA 300
C       INFIL        MORE FILE FORTRAN UNIT NO.   15            USER        GTDA 305
C                     MUST CORRESP WITH JCL FOR  @FCF--SEE JCL BELOW        GTDA 310
C       IDNUL        VALUE FOR NULL INDICATOR     10            USER        GTDA 315
C       VALNUL       VALUE FOR NULL QUANTITY      1.E-15        USER        GTDA 320
C                                                                           GTDA 325
C                                                                           GTDA 330
C                                                                           GTDA 335
C                                                                           GTDA 340
C* SOME  VARIABLES ALLOW THE USER TO GET THE STATION ID IN A FORMAT         GTDA 345
C     OTHER THAN IN THE FOLLOWING DEFAULT FORMAT:                           GTDA 350
C      NSCODE(1) AND (SCODE(2) HAVE THE PRIMARY STATION NUMBER              GTDA 355
C       IN 2A4 FORMAT,91X IN FROM THE BEGINNING OF THE FIRST STATION        GTDA 360
C      RECORD.    THE SECONDARY NUMBER IS READ IN IN A4,A3,A2 FORMAT,       GTDA 365
C       C AFTER SKIPPING 34 SPACES.                                         GTDA 370
C       KODSIZ IS THE NUMBER OF WORDS DESIRED, AND MAY BE AS MUCH AS 16.    GTDA 375
C                                                                           GTDA 380
C       THE KFMT ARRAY MUST CONTAIN THE FORMAT SPECIFICATIONS DESIRED.      GTDA 385
C                                                                           GTDA 390
C***  J C L  ***  AND MINIMUM COMPLEXITY USAGE OF GETDAT                    GTDA 395
C                                                                           GTDA 400
C      //MJFRT  JOB  (LV03,35,1,1),'USE-GETDAT',MSGLEVEL=(0,0)              GTDA 405
C      //RETREV EXEC WUDIST                                                 GTDA 410
C      //DIST.CARDFD DD *                                                   GTDA 415
C      PGM=RET,A=1119URPF,MORE=3,RMT=35,                                   GTDA 420
C      B,S=002000130,S=PMN005,                                             GTDA 425
C      P=85520,P=85514,BD=720101,ED=721231,                                GTDA 430
C      PRT=NO,                                                             GTDA 435
C      /*                                                                  GTDA 440
C      //USE  EXEC FORTGCLG                                                GTDA 445
C      //FORT. SYSIN DD *                                                  GTDA 450
C      C  THIS USER PROGRAM USES GETDAT                                    GTDA 455
C      C                                                                   GTDA 460
C           COMMON /GVALUE/VALUE(10)                                       GTDA 465
C           COMMON /GPARM /NDFIL,NUMRET                                    GTDA 470
C      C FIRST CALL PRECEDED BY SETTING NUMBER OF PARAMS RETRIEVED          GTDA 475
C           NUMRET=2                                                       GTDA 480
C         5 CALL GETDAT                                                    GTDA 485
C           IF(NDFIL.NE.0)GO TO 90                                         GTDA 490
C           PRINT 7,(VALUE(I),I=1,NUMRET)                                  GTDA 495
C         7 FORMAT(2X,2E15.2)                                              GTDA 500
C           GO TO 5                                                        GTDA 505
C      C EOF                                                               GTDA 510
C        90 STOP                                                           GTDA 515
C           END                                                            GTDA 520
C      //LKED.ADDLIB DD DSN=CNLV03.MJF.GETDAT,UNIT=3330,                   GTDA 525
C      //  VOL=SER=TS0004,DISP=SHR                                         GTDA 530
C      //LKED.SYSIN DD *                                                   GTDA 535
C        LIBRARY ADDLIB(GETDAT)                                            GTDA 540
C      //GO.FT15F001 DD DSN=&FCF,DISP=OLD                                  GTDA 545
C                                                                           GTDA 550
C*****                                                                      GTDA 555
C                                                                           GTDA 560
      SUBROUTINE GETDAT                                                     GTDA 565
      DIMENSION ACODE(2)                                                    GTDA 570
      DIMENSION PNAME1(50,4),PNAME2(50,4), STNID(9,10)                      GTDA 575
      COMMON /GVALUE/VALUE(50)                                              GTDA 580
      COMMON /GNULL/NULL(50)                                                GTDA 585
      COMMON/GCODE/NSCODE(16),NDATE(3),NTIME(2)                             GTDA 590
      COMMON /GKNTR/KNTSTA,KNTREC,KNTNUL(50)                                GTDA 595
```

Figure 4 (Continued)

```
      COMMON /GPARM/NUMORE,NUMRET,MORE,INFIL,IDNUL,VALNUL                GTDA 600
      COMMON /GFORM/KODSIZ,KFMT(10)                                      GTDA 605
      COMMON /GECHO/ ECHO                                                GTDA 610
      COMMON /GDEEP/DEPTH
      DATA KALL01/0/                                                     GTDA 615
            DATA VLNUL/Z2F4B8ED0/
C TEST-FLAG FOR DIAGNOSTIC OUTPUT                                        GTDA 620
      KTEST=0                                                            GTDA 625
      KTEST=1                                                            GTDA 630
C  IF FIRST CALL DONE, GO GET DATA                                       GTDA 635
C                                                                        GTDA 640
      IF(KALL01.NE.0)GO TO 400                                           GTDA 645
      KALL01=1                                                           GTDA 650
C                                                                        GTDA 655
C CHECK PARAMETERS THAT COULD ALLOW USER SELF-DAMAGE                     GTDA 660
      KODSTP=0                                                           GTDA 665
      IF(NUMRET.GT.10.AND.MORE.NE.4)KODSTP=1                             GTDA 670
      IF(KODSIZ.GT.16)KODSTP=2                                           GTDA 675
      IF(KODSTP.EQ.0)GO TO 30                                            GTDA 680
      PRINT 28,KODSTP                                                    GTDA 685
   28 FORMAT(' ERROR IN PARAMETER OVERRIDDEN,NUMBER=',I6)                GTDA 690
      STOP 28                                                            GTDA 695
C                                                                        GTDA 700
C                                                                        GTDA 705
   30 CONTINUE                                                           GTDA 710
C                                                                        GTDA 715
C                                                                        GTDA 720
      PRINT 35,NUMRET,MORE,INFIL,IDNUL,VALNUL,                           GTDA 725
     X  KODSIZ,KFMT                                                      GTDA 730
   35 FORMAT('0NUM.PARS,MORE VALUE,INFIL,NULL INDIC,NULL VAL,            GTDA 735
     X 'KODSIZ,FORMAT'/ 1X,                                              GTDA 740
     X  4I6,E15.3,I6, 10A4 )                                             GTDA 745
C  PARAMETER  RECORDS                                                    GTDA 750
      DO 205 K=1,5                                                       GTDA 755
      ILOW=10*(K-1)+1                                                    GTDA 760
      IHIGH= 10*K                                                        GTDA 765
C                                                                        GTDA 770
      DO 200 J=1,4                                                       GTDA 775
      READ(INFIL,201,END=599,ERR=799)                                    GTDA 780
     X                        (PNAME1(I,J),PNAME2(I,J),I=ILOW,IHIGH)    GTDA 785
      IF(ECHO.EQ.1)PRINT 202,                                            GTDA 790
     X  (PNAME1(I,J),PNAME2(I,J),I=ILOW,IHIGH)                           GTDA 795
  200 CONTINUE                                                           GTDA 800
      IF(MORE.NE.4)GO TO 210                                             GTDA 805
  205 CONTINUE                                                           GTDA 810
  201 FORMAT(45X,10(2A4,2X) )                                            GTDA 815
  202 FORMAT(' ECHO 1:',10(2A4,2X))                                      GTDA 820
  210 CONTINUE                                                           GTDA 825
C                                                                        GTDA 830
C  DATE RECORD                                                           GTDA 835
C     DUMMY ALL 99 99 99                                                 GTDA 840
C                                                                        GTDA 845
      READ(INFIL,300)DATE                                                GTDA 850
  300 FORMAT(25X,3I2)                                                    GTDA 855
      IF(ECHO.EQ.1)PRINT 301,DATE                                        GTDA 860
  301 FORMAT(' ECHO 2:',3I3)                                             GTDA 865
C                                                                        GTDA 870
C  GO HERE AFTER FIRST CALL                                              GTDA 875
C                                                                        GTDA 880
C  STATION ID RECORDS                                                    GTDA 885
C                                                                        GTDA 890
```

Figure 4 (Continued)

```
  302 CONTINUE                                                          GTDA 895
      READ(INFIL,KFMT,END=699,ERR=799)(NSCODE(IJ),IJ=1,KODSIZ)          GTDA 900
      IF(ECHO.EQ.1)PRINT  305,(NSCODE(IJ),IJ=1,KODSIZ)                  GTDA 905
      DO 303 I=2,9                                                      GTDA 910
      READ(INFIL,304,END=599,ERR=799)                                   GTDA 915
     X                          (STNID(I,J),J=1,10)                    GTDA 920
      IF(ECHO.EQ.1)PRINT 305, (STNID(I,J),J=1,10)                       GTDA 925
  303 CONTINUE                                                          GTDA 930
      KNTSTA=KNTSTA+1                                                   GTDA 935
  304 FORMAT(90X,10A4)                                                  GTDA 940
  305 FORMAT(' ECHO 3:',10A4)                                           GTDA 945
  400 CONTINUE                                                          GTDA 950
C
C  IF MORE =4 READ ALL 50 VALUE ITEMS AND DEPTH, OTHERWISE,
C   JUST READ THE NUMBER OF PARAMS RETRIEVED,  (NUMRET)
C
      IF(MORE.NE.4)GO TO 405
      READ(INFIL,410,END=699,ERR=799)NDATE,NTIME,VALUE,DEPTH
  410 FORMAT(25X,5I2,50A4,64X,F5.0)
      GO TO 415
  405 CONTINUE
      READ(INFIL, 410,END=699,ERR=799)NDATE,NTIME,(VALUE(I),I=1,NUMRET) GTDA 955
  415 CONTINUE
      NUMREC=NUMREC+1                                                   GTDA 965
      IF(ECHO.EQ.1)PRINT 410,NDATE,NTIME,(VALUE(I),I=1,NUMRET)          GTDA 970
  403 FORMAT('0ECHO 4')                                                 GTDA 975
      IF(ECHO.EQ.1.AND.MORE.EQ.4)PRINT 411 ,DEPTH
  411 FORMAT(1X,'DEPTH:',F10.3)
      IF(NDATE(1).EQ.99)GO TO 302                                       GTDA 980
C                                                                       GTDA 985
C ADJUST VALUES                                                         GTDA 990
C                                                                       GTDA 995
      DO 433 I=1,NUMRET                                                 GTDA1000
        NULL(I)=0                                                       GTDA1005
      IF(VALUE(I).NE.VLNUL)GO TO 433                                    GTDA1010
      VALUE(I)=VALNUL                                                   GTDA1015
      NULL(I) =IDNUL                                                    GTDA1020
      KNTNUL(I)=KNTNUL(I)+1                                             GTDA1025
  433 CONTINUE                                                          GTDA1030
C                                                                       GTDA1035
C     UP RECORD COUNT AND RETURN                                        GTDA1040
C                                                                       GTDA1045
      KNTREC=KNTREC+1                                                   GTDA1050
      RETURN                                                            GTDA1055
C                                                                       GTDA1060
C  EOF DURING PARAMETER ID RECORDS                                      GTDA1065
C                                                                       GTDA1070
  599 NDFIL=2                                                           GTDA1075
      PRINT 600,KNTSTA,KNTREC                                           GTDA1080
      STOP 599                                                          GTDA1085
  600 FORMAT('1EOF DURING PARAM OR STATION ID READING,STA/REC=',2I7)    GTDA1090
C                                                                       GTDA1095
C  EOF DURING DATA RECORD                                               GTDA1100
C                                                                       GTDA1105
  699 NOMORE=1                                                          GTDA1110
      RETURN                                                            GTDA1115
C                                                                       GTDA1120
C  ERROR DURING READING                                                 GTDA1125
C                                                                       GTDA1130
  799 NOMORE=2                                                          GTDA1135
      PRINT 800,KNTSTA,KNTREC                                           GTDA1140
```

Figure 4 (Continued)

```
      STOP 799                                                          GTDA1145
  800 FORMAT('1ERROR DURING READING')                                   GTDA1150
      END                                                               GTDA1155
      BLOCK DATA                                                        GTDA1160
      COMMON /GNULL/NULL(50)                                            GTDA1165
      COMMON /GKNTR/KNTSTA,KNTREC,KNTNUL(50)                            GTDA1170
      COMMON /GPARM/ NOMORE,NUMRET,MORE,INFIL,IDNUL,VALNUL              GTDA1175
      COMMON /GFORM/KODSIZ,KFMT(10)                                     GTDA1180
      COMMON /GECHO/ ECHO                                               GTDA1185
      DATA KNTSTA,KNTREC,KNTNUL/ 52*0/                                  GTDA1190
      DATA NOMORE,NUMRET,MORE,INFIL,IDNUL,VALNUL/ 0,10,3,15,10,1,E-15/  GTDA1195
      DATA KODSIZ,KFMT/5,4H(91X,4H,2A4,4H,34X,4H,A4,,4HA3,A,4H2)   /    GTDA1200
      DATA NULL/50*0/                                                   GTDA1205
      DATA ECHO/0,/                                                     GTDA1210
      END                                                               GTDA1215
```

Figure 4 (Continued)

TWO DATA STORAGE PROGRAMS FOR STORET

By Kenneth V. Byram

SCAN - A PREEDITING PROGRAM FOR STORET DATA

Purpose

This program and its two cataloged procedures were written to enable a STORET "storer" to make arithmetic checks on his data before it is actually stored in the STORET data base. The program also produces a fairly tabular listing of all of the data. It is hoped that by using this procedure before entry of data into the system, a storer can eliminate some of the more obvious kinds of errors, especially those that are caused by clerical mistakes stemming from the nontabular forms of most STORET water quality input. A pH value of 75, for example, is particularly embarrassing and obvious when it appears neatly tabulated on a STORET retrieval; such an occurrence casts doubt on all of the user's data. It is difficult to eliminate those kinds of errors, particularly if the data has originated on a long string of DIP input, where the 75 is adjacent to parameters where 75 is a legitimate value, or from a STORET standard format card with the value coded as "00400750012."

What It Does

This program, for one thing, produces tabular listings of data, so that a pH of 75 should be somewhat more obvious. Under user control, it also makes "standards" type checks on the data, checking against a user-input set of maximum and minimum values by parameter. For example, a minimum of 2 and a maximum of 12 might be coded for pH. Also under user control, it makes interparameter comparisons. A measurement of total rivets, for example, must logically always be greater than, or equal to, a measurement of dissolved rivets on the same sample. This comparison would flag an occurrence of the values 10 for total rivets, and 15 for dissolved rivets. The program additionally makes checks for parameters which are duplicated within the input for that run. It will not detect duplications of data previously stored in the STORET system.

What You Provide as Input to the Scan

To Get the Tabulation - No input need be provided. The program will always tabulate whatever parameters are there.

To Make a Within Limits Comparison - The input provided is a parameter code, the minimum value for that parameter, and the maximum value. The parameter code goes in columns 1-5, right justified. The minimum goes in columns 11-20, with decimal point explicit, with the maximum similarly coded in columns 21-30. Columns 6-10 should be blank. (A maximum of 100 such checks may be input.)

To Make an Interparameter Comparison - The input provided is two parameter codes, and the resultant test is that the value for the first is greater than, or equal to, the value for the second. An error message is printed if the first value is less than the second. The first parameter code goes right justified in columns 1-5, and the second in columns 6-10. Other columns are not read by the program. (A maximum of 100 such checks may be input. These are in addition to the 100 above.)

To Sort or Not to Sort - There are two cataloged procedures, one that leaves the input data as is, and the other that sorts it by agency, station, date and time. In a sense, the sort is a disadvantage because it reorders the input, listing errors detected by the program in an order different from that in the input deck. The sort is an advantage in viewing the tabulation, particularly if the input is in somewhat random order. It is also advantageous if there are possibilities for interparameter comparisons, or if there are duplicate parameters at different places in the input deck. Without the sort, the comparison would not be made.

Using the Program

The program executes a simulated STORET storage run, bypassing the data sets that are accumulated for weekly mass storage runs. Thus, normal errors of

"station not presently in system," "invalid parameter," etc., will be printed out. The program uses 190K of core, and will take only slightly longer than a normal storage run.

The deck setup is:

```
// (job card)
// EXEC SCANWQJ (or SCANWQJS, for the sort)
// DIST. CARDF DD *
   followed by a standard STORET storage deck:
      ?START
      ?01,02,03 or 04 . . . etc.
// SCAN.SYSIN DD *
```

Scan program cards are described above. Cards are unnecessary if no limits or interparameter comparison checks are to be made. Limit and interparameter comparison cards may be intermixed.

Notice that running either procedure will not store the data in STORET. One would usually run a scan, correct the errors, run another scan to check the corrections, and then run a STWQJ to actually store the data.

Modifying the Program

For those interested in modifying the FORTRAN program to add checks of their own, the source deck is available on &CNA907.KVB.SCAN on TS0004. Once the modification is ready, JCL for making the modified STORET input run, compiling the program, and running it is available on &CNA907.KVB.SCAN.JCL. The sorted version is &CNA907.KVB.SCAN.JCLS. The existing FORTRAN program uses an assembler language subroutine which will be assembled in the run. The JCL must be modified to reference the program, the STORET input, and the program input data.

For more information, call Ken Byram, at FTS 503-752-4385, or commercial at 503-752-4211 ext 385. This writeup will be maintained and modified if program changes are made. It is available on &CNA907.KVB.-SCAN.DOCUMENT.

SUBROUTINE STORE

Purpose

STORE is a FORTRAN subroutine which prepares a complete card-image deck for data storage in STORET, including header cards, agency card and data cards, when called from a calculation program.

Most STORET users input their water quality data with either the DIP format or the "standard" (five/parameters per card) format. Neither of these formats is particularly easy to produce from a running FORTRAN program, and the user who wishes to use a program to compute water quality data from intermediate factors has the option of recoding the data after it is listed by the program or manipulating his FORTRAN to produce the required output. This subroutine allows the programmer to put date into STORET card format directly from his program by making calls to the subroutine, which then does all necessary left zero filling and data formatting. The deck produced by the subroutine contains the necessary header cards, so with addition of the proper JCL, the deck may be entered in the same job. As an example, a user new to STORET may want to calculate and enter chlorophyll data. The data manipulation requires that 6 absorbance series of formulae which check for unrealistic volumes, calculate concentrations, round them off, and check for precision. Without having to learn all the restrictions of the STORET system, the user can write his program, create entry records, add the JCL, and enter data into STORET.

HOW TO USE STORE

There are five calls made from the main program to store data in the STORET storage deck.

1. Call Agency(I)

This call must be made at the beginning of the program. It writes the header cards and copies the information for the agency card from an array to the storage deck. This call is made only once and must be made before any other calls to the subroutine.

(I) is the integer array name into which the agency card has been copied. The expected format is (20A4).

2. Call STOOBS (N,S,ID,IDATE,ITIME)

This call gives the subroutine the station, date, and time for the entry of any values following the call. It is called once for each time the station, date, or time changes.

(N) is the number of characters in the station code, up to the maximum allowed, 15 characters.

(S) is the array in which the station code is stored. It may be from 1 to 4 words, depending on the number of characters specified.

(ID) is the depth. It is an integer value used as a sample locating parameter. If it is negative, depth will be considered as not present as a sample locating parameter.

IDATE is the 6 digit integer date in the form YYMMDD where YY is the year, MM is the month, and DD is the day of the month.

ITIME is the 4 digit integer time, using a 24 hour clock. If ITIME is negative, it will be considered as not recorded.

3. Call STOCDE (IP,V,R)

This call transfers to the subroutine the data which is to be stored. The previous STOOBS call provides the station, date, and time for storage. When five parameters have been stored, it dumps the card into the STORET storage deck.

(IP) is the STORET parameter code. It is an integer value, which the subroutine provides with any necessary leading zeros.

(V) is the floating point value to be stored for the specified parameter code. No more than 4 significant digits are accepted by the subroutine; so if there are more than 4, the subroutine rounds the value off. Values must be in the range E-10 to E+9. Zero is not allowed.

(R) is the one character mark. A list of remark codes can be found on page 7-7 of the STORET manual. Usually this would be a blank (1H); but, for example, if the user wanted to indicate that an answer was an estimate, he would use 1H0.

4. Call STODMP

Up to five parameters may be stored on each storage card. STOCDE calls transfer the information on each parameter to the subroutine. When all parametric values have been stored for a given station, STODMP is called. It must be called once for each STOOBS call. STODMP dumps any incomplete (less than 5 parameters) cards and reinitializes the subroutine for the next STOOBS call.

5. Call STODEL (IP)

STODEL creates a record to delete one parameter for a given station-date-time. As with the STOCDE call, the station, date, and time are provided by the previous STOOBS call.

(IP) is the STORET parameter code in integer form.

RUNNING A PROGRAM USING STORE

A simple way of using STORE is to copy it onto the end of your program and use the following WYLBUR commands:

```
?USE (your file) ON (your volume)
?COPY ALL TO END FROM &CNA907.KVB.
   STORE ON TS0004
?SAVE (your name) ON (your volume).
```

When you execute the program, you could use the following JCL to create a SYSOUT DATA set with STORET card images for checking:

```
// job card
// EXEC EDSIN,NAME=(your filename),DISK=
   (your volume)
// EXEC FORTGCLD
// FORT.SYSIN DD DSN=&INPUT,DISP=(OLD,
   DELETE)
// GO.SYSIN
   your cards
// GO.FT08F001 DD SYSOUT=A.
```

If you want to make a STORET storage run directly with [illegible] program, replace the //GO.FT08... card with the following:

```
// GO.FT08F001 DD DSN=&STORET, DISP=
   (NEW,PASS),UNIT=SYSDA,
// SPACE=(TRK, (10,10), RLSE),DCB=(BLKSIZE=
   3520,LRECL=80,RECFM=FB)
// EXEC STWQJ
// DIST.CARDF DD DSN=&STORET,DISP=(OLD,
   DELETE).
```

For more information, call Ken Byram, at FTS 503-752-4385, or commercial at 503-752-4211 ext 385. This writeup will be maintained and modified if program changes are made. It is available on &CNA907.KVB.-STORE DOCUMENT ON TS0004.

THE EPA SCIENTIFIC APPLICATIONS SOFTWARE STUDY

By Elijah L. Poole

BACKGROUND

The Environmental Protection Agency (EPA) uses several computer facilities in fulfilling its functions. These facilities encompass minicomputers as well as intermediate and large-scale digital computers. Many of the small computer systems are special purpose computers that service a single laboratory. However, most of the large computer systems have on-line remote terminals and serve a large community of users. The scientific applications software systems implemented on these computers are both numerous and varied. It is the large-scale computers that this study is designed to cover.

There have been several recent studies of EPA computing systems requirements performed under contract to EPA. One of these, performed by Computer Sciences Corporation (CSC), resulted in an inventory of EPA Information and Administrative Systems. It excluded, however, systems which were considered "purely scientific."

Another study, carried out by the General Electric Corporation (GE), surveyed all ADP requirements in the Agency with a view toward determining the need for large-scale computing centers.

It was decided that a scientific applications software study should be performed to help in fulfilling the Agency's needs in this area. So a task was negotiated with Systems Architects, Incorporated to provide level of effort support for ten months beginning September 9, 1974.

SCOPE

This task requires that a survey be conducted of the scientific software packages implemented on the major computer systems utilized by EPA. The computer systems to be included in the survey are specified below:

Organization	Computer Systems
Washington Metropolitan Area	
Optimum Systems Incorporated	IBM 370/155, 158
National Institutes of Health	IBM 370/165 PDP-10
Research Triangle Park, N.C.	
NERC-RTP	UNIVAC 1110 IBM 360/50

METHODOLOGY

The approach and basic functions of the study are as follows:

1. Review EPA requirements as documented in the CSC and GE studies and as augmented by interviews and other available documentation. Make recommendations to acquire additional scientific applications software, or to take any other measures that would improve the Agency's use of this type of software.

2. Conduct a survey of what is presently available on the computer systems described above in terms of scientific applications software packages such as mathematical packages, statistical packages, modeling systems (GPSS, GASP, CSMP, etc.), biomedical computer programs, and plotting routines. Poll users to determine if there are additional software needs and implement such software where feasible.

3. Develop a user's manual that would describe the capabilities, advantages and disadvantages of the available software, give instructions on how to use it, and provide criteria to determine which software system should be used for a given application.

4. Identify and document user training needs for scientific applications software. Provide a recommended schedule for training and provide training courses to the extent possible.

BIOMED, SAS, AND OTHER TOOLS OF THE STATISTICIAN

By Robert Kinnison, Ph.D.

STATISTICAL PACKAGES

Presently there are two widely used statistical packaged systems: the Statistical Analysis System (SAS) and the Biomedical Analysis Programs (BIOMED). These analytical tools, differing essentially only in their details and emphasis, may be considered to be in competition for the user community. A historical perspective is a convenient means of comparison. The early series of BIOMED was the first widely disseminated and accepted program package available and was first released in 1965. An immediately apparent handicap for the experienced user (but definitely not for the student or neophyte user) was the rigidly structured fixed format specification of input. The first widely known attempt to relax this structure was the Statistical Package for the Social Sciences (SPSS) which had an English language type input specification language. Designed for data description only, SPSS suffered from a lack of completely free field input and a very limited statistical capability.

SAS was the next major improvement in widely accepted packages. It included a basically free field input, time sharing orientation, excellent data editing capability and an expanded statistical capability. Not to be outdone, the BIOMED programmers also developed a soon to be released update of their package, called BMDP, that has many features not found in previously released systems. (This package has actually been selectively released for field evaluation.) This BMDP series of programs has free field input, a data editing capacity that is equivalent to SAS (although SAS is easier to use in this respect) and a much more extensive statistical capability than any previously released packages. It also has a wide range of output options for data description. Its basic capabilities that extend beyond those of other packages are: time series analysis (Fourier analysis, autocorrelation etc.), cluster and classification procedures, multivariate analysis procedures (beyond multivariate analysis of variance), contingency table analysis, many nonparametric tests, nonlinear regression, automatic hypothesis testing in the general linear hypothesis, life tables, survival rates, and biological assay.

There are several other packages that are available but not so widely accepted. Most universities have their own unique systems, available only on their own computer containing most of the more commonly used features of SAS and BMD. The National Bureau of Standards has released a program series called OMNITAB, which is oriented to data screening and is weak on statistical analysis in comparison to SAS and BIOMED. Additionally there is a wide variety of subroutines available with the capabilities of the program packages, but they require computer programming ability since these subroutines must be used in conjunction with user-provided input and output programs. Such subroutines are found in IBM's Scientific Subroutine Programs (SSP), a commercial update of SSP called International Mathematical and Statistical Library (IMSL), CDC's Statistical Subroutine Library, and Union Carbide's Subroutine Library. Because of the expertise necessary to use these subroutines, they can be recommended only for special purpose use.

Because of the close similarity of SAS and BIOMED and their extensive capabilities, they can be recommended over all others. The choice between them can be based on personal preference. For the nonstatistically trained person, SAS is certainly adequate and easier to use. For the professional statistician, the extra statistical capability of BIOMED favors its use. However, an overriding consideration would be the existence of either package on a readily available or convenient computer system.

UTILITY ROUTINES AND OTHER TOOLS

Any statistician, of any level, will quickly find himself in situations where some computer programming is highly desirable for the execution of his duties. He will invariably start with a programmable calculator and soon graduate to a generalized scientific computer language such as FORTRAN. At this point a personal library of subroutines, functions, and algorithms becomes an almost essential tool. In addition to the rather comprehensive libraries, such as SSP (mentioned previously), the following is this author's selection of useful but not well-known programs. Further details, original sources, or program copies and general comments will be gladly provided to anyone contacting or writing this author.

It is most useful to have the capability of finding exact probability levels without consulting large statistical tables of the normal, t-test, chi-square, or F-test statistics. Special and efficient subroutines are available for each of these common statistics, and also for less common distributions such as the Gamma function. For

rather infrequent use, a convenient relationship exists that permits the computation of all common probability levels from only a subroutine for the probability of the F-test. The following equalities use commonly accepted notation for degrees of freedom and significance level:

$$t_v(a/2) = \left(F_{1,v}(a)\right)^{1/2}$$

$$z(a) = t_\infty(a)$$

$$X_v^2(a) = v\,F_{v,\infty}(a)$$

The subroutine called FCDF takes the F-value and returns the corresponding probability level a. It was originally implemented in FORTRAN by the BIOMED group at the University of California, Los Angeles (UCLA). It is available separately only as a listing and is only 56 cards long. For practical purposes, infinite degrees of freedom may be computationally specified as one to two thousand for four digits of accuracy and several hundred for three digits of accuracy. This routine uses a highly accurate numerical approximation and should not be confused with several available statistical approximations with accuracy only within 10 percent.

A remarkably common and unrecognized problem is one of roundoff error when calculating standard deviations and correlations. This has become important recently with the proliferation of hand and programmable calculators. Most elementary statistics courses are taught from text based on computation with mechanical calculators. A mathematical identity commonly used computationally in order to enter the data only once into the calculator is:

$$\sum_i \left(x_i - \bar{x}\right)^2 = \sum_i x_i^2 - \left(\sum_i x_i\right)^2 .$$

What is not generally known, except among professional statisticians, is that this identity is not generally applicable to electronic computing equipment. A problem arises because two very large sums may be accumulated until their statistically significant parts are lost to roundoff or truncation errors which occur automatically in electronic calculators but not in mechanical ones. An algorithm is available that has the accuracy of first calculating the mean, then repeating the data to find deviations; the data need be entered only once:

$$SS_n = SS_{n-1} + \frac{(n-1)}{n}\left(x_n - \bar{x}_{n-1}\right)^2$$

where SS_n is the sum of squares for a sample size of n, X_n is the nth data point and $\bar{X}_{n-1}$ is the average calculated from n-1 data points, and SS_o and $\bar{X}_o$ both equal zero.

One of the more commonly used classes of statistical computation is that of contingency table analysis. Computational algorithms and programs are readily available for two-way tables with equal sample size. When more than two categories of classification are desired, when data are missing, or when sample sizes within the categories are different, these algorithms are not valid. The common practice of dividing the data into many simpler contingency tables results in invalid statistics in the same sense that the common practice of computing a t-test or a correlation for all possible combinations of variables is wrong. An elegant multidimensional contingency table analysis theory has recently been developed in the statistical literature; it allows for analysis of variance type hypotheses and factors to be statistically tested and is valid for any contingency table.* An implementation of these methods is available from Dr. Leo Goodman of the University of Chicago. It must be emphasized that the use of these programs, while essential for many contingency table problems, is a highly sophisticated statistical procedure that is full of hidden assumptions and pitfalls. Thus they should only be used under the guidance of a qualified statistician.

A very commonly used class of procedures in epidemiology is classification of cases based on the logistic transformation. This type of analysis finds the probability of a response, such as having a disease, based on a polynomial type combination of the possible underlying contributing factors. It has been found to be appropriate with a wide number of types of measurement factors used simultaneously, from accurately measured quantitative variables to gross and dichotomous classifications; it has proven highly effective in predicting the average response of populations and subpopulations. Several prominent names are associated with slight variations of the algorithm and the several corresponding implementations into computer programs. Walker and Duncan publish mainly in the American statistical literature, Armitage in the English medical literature, Meshalkin in

*The theory behind this method is reviewed in R.L. Plackett, "Multidimensional Contingency Tables: A Survey of Models and Methods," *Institute International de Statistique Bulletin* (Rome, 1969), Vol. 43, p. 133 ff.

World Health Organization reports, and Cornfield in the American medical journals.

A somewhat specialized but highly efficient analysis of the variance tool is available from Elliot Cramer, Acting Director of the Psychometric Laboratory at the University of North Carolina, Chapel Hill. This program package, called MANOVA, is a highly sophisticated multivariate analysis of the variance system. It was not discussed with SAS and BIOMED because of its limited scope and acceptability. It should be noted that in everyday data analysis, multivariate analysis of variance problems is almost unknown. The strength of this package is its ability to handle unequal sample sizes and to automatically formulate the more common hypotheses for analysis. It should also be noted that this program can handle any analysis of variance, analysis of covariance or regression problem (except stepwise regression) as a subset of its overall capability.

Another common statistical problem is that of selecting those regression variables which truly statistically explain the observed effects and of rejecting those variables which add only noise to the system. The well-known procedures for this selection are classified as the "Stepwise Regression Algorithms." Early users of these algorithms noted that their regression results depended upon peculiarities of the data and upon the details of the algorithm implementation. The most widely tested algorithm for avoiding these problems is to run a separate regression on each and every combination of regression variables and choose the one with the smallest residual mean square error; this, of course, takes excessive computer time for most regression problems. An efficient compromise was reported by L. R. LaMotte and Hocking.* The implementation of this algorithm by LaMotte is called program SELECT.

A problem in numerical analysis that is encountered by statisticians whenever a regression or analysis of variance procedure is implemented, is that of matrix inversion. The commonly-used techniques, such as Gauss-Jordan or pivotal condensation, frequently give the wrong answer whenever the design matrix is ill-conditioned (nearly singular or has a determinant close to zero), or whenever the specification of restrictions on the parameters is incorrectly done or poorly conceived. A very general mathematical tool can be used in this statistical problem because of its inherent numerical stability. Known as the "generalized matrix universe," it is a concept that specifies an inversion of any matrix or array regardless of its properties. This concept is not restricted to square matrices or matrices of full rank which are typical of statistical computation; however, when applied to these statistical computations, it offers a stable and efficient algorithm that is not affected by the conditioning qualities of the matrix. Furthermore, in cases where the design matrix is truly singular (the determinant is identically zero), it is known that there is an infinite number of answers, but the generalized matrix inverse gives that one answer that is of minimum norm; that is, the one in which the sum of the squares of the parameter values is minimum. An implementation of this algorithm, as a subroutine, is available from the author as program MATGIV.

Nonlinear regression algorithms have received extensive use in some laboratories for calculating such quantities as biological or radiological half-lives. Such problems are inherently based on systems of differential equations which are handled in the regression programs as sums of exponential terms. From the standpoint of numerical analyses, sums of exponentials are notoriously prone to blow up computationally in the commonly used algorithms such as the Gauss (used in BIOMED and SAS), Marquardt's gradient method, and the less well-known available ones such as SPIRAL. It should be noted that for nonlinear regressions, not of the exponential form, these algorithms are very satisfactory. A recently discovered algorithm helps substantially overcome the problems associated with sums of exponentials; this class of algorithms and theory is known as Variable Projection methods. The name and methods are so new that this nomenclature is not firmly established and some authors may refer to it by other names.† This technique divides the regression parameters into two classes, linear parameters (in which the first derivatives of the regression function are independent of any of these parameters) and nonlinear parameters. Through some special transformations of the mathematical sample space, the algorithm is able to find unbiased estimates of the nonlinear parameters. After these estimates are known, the remaining parameters are easily found by simple linear regression. A distinct computational advantage is gained by working with small subsets of parameters. Even though this algorithm will converge to a valid

*L.R. La Motte and Hocking, " Computational Efficiency in the Selection of Linear Regression Variables, "*Technometrics*, Vol. 12, No. 1 (February 1970).

†For reference, see Guttinan, Pereya, and Scolnik, "Least Squares Estimation for a Class of Nonlinear Models," *Technometrics*, Vol. 12, No. 2 (May 1973), p. 209 ff.

answer in situations for which the more common nonlinear programs breakdown, it should be emphasized that these situations are complex mathematical and statistical phenomena. Therefore, these powerful new algorithms are advised for use only by persons with a strong statistical background who are familiar with the implications of the algorithms. A well-programmed implementation is available at Stanford University Computer Center; it is called VARPRO and was developed by Professor Gene Golub of the Computer Sciences Department. This author has a slightly modified version of VARPRO that is easier to use but not so computationally efficient.

Perhaps the most common simulation or modeling algorithm used for environmental transport studies is systems of linear differential equations. This type of model is also extensively used in chemical kinetic studies and in biological metabolism investigations. The previously mentioned nonlinear least squares algorithms are not capable of estimating the coefficients of these equations from sample data unless the systems of differential equations are reformulated as systems of integral equations; this reformulation is an essentially impossible task. However, because of the importance of this problem in drug metabolism, the Upjohn Company has developed a program package that can estimate the parameters of these differential equations from data. The technique used can best be described as brute force numerical integration and differentiation, yet it does work well and has been extensively tested. Their currently available package of programs also has general nonlinear regression capacity and aids in determining confidence limits on the nonlinear parameters and algorithms for determining optimum sampling designs. The contact for this package is Dr. C. M. Metzler, Upjohn Company, Kalamazoo, Michigan.

A commonly desired algorithm is one to find the maximum (or minimum) of some algebraic function; the most obvious example of its use is the need to find the optimum conditions for environmental sampling. Classically this type of problem has been divided into problems of optimizing economic considerations, for which the techniques of operations research, linear programming, and nonlinear programming were developed, and of mathematical optimization with a large number of algorithms. The most well-known and best accepted algorithm is the Davidon-Fletcher-Powell algorithm. Recent work has shown that with some ingenious implementation, the operations research algorithms are really subsets of the mathematical optimization. A practical disadvantage of the Davidon-Fletcher-Powell algorithm, which is available in SSP and IMSL, is that it requires the user to supply the partial derivatives of the function being optimized with respect to the parameters; this is tedious for many functions and always invites algebraic mistakes by the programmer. Recently a group of programs has been published which accurately approximates these derivatives using a numeric technique known as the "principle axis method." A good implementation of this algorithm is called PRAXIS and is a subroutine available from the Department of Computer Sciences of Stanford University.* Articles on this subject frequently appear in *The Computer Journal*, and in *The Journal of the Society of Industrial and Applied Mathematics (SIAM Journal)*.

For the solution of operations research problems, many program packages are available but not well publicized. The more recent and better packages are capable of mathematical programming and nonlinear programming as well as the more classical linear programming solutions. Because of this increased capacity, they are also more difficult to use; therefore, it can be recommended that the beginner with a simple linear programming problem use one of the FORTRAN programs listed in many elementary operations research textbooks. A highly sophisticated generalized program is called Sequential Unconstrained Minimization Technique for Nonlinear Programming (SUMT) and is available with documentation at a cost of $100 from the Research Analysis Corporation of McLean, Virginia. Because this program uses the Davidon-Fletcher-Powell algorithm internally, it can be used for any optimization or minimization problem, including nonlinear regression and optimum sampling schemes. The only restriction is that the user must be able to express his problem in a FORTRAN subroutine.

A variety of generalized simulation programs are available and have been used for many years by the engineering profession. These programs had their origin in the extensive use of analog computers by the aircraft industry. With the general availability of scientific digital computers, these engineers switched to the more flexible digital computers. Because of the historic association with analog computers, the software packages developed required that the problem be formulated in terms applicable to analog computers; the situation has not changed over the years because this formulation is a rather natural language for structural engineers and chemical

*For reference, see Richard C. Brent, *Algorithms for Finding Zeros and Extrema of Functions Without Calculating Derivatives*, Chap. 7.

kinetics. It is fortuitous that it is also a natural language for environmental pollutant transport models. Currently there is a remarkable standardization of these program packages, at least as far as the user is concerned in setting up his problem (internally there are marked differences). All program packages are based on a standard established and maintained by a professional group called Simulation Councils and are called Continuous Systems Simulation Languages (CSSL). The IBM implementation is called Continuous Systems Modeling Program (CSMP) and is available on all of their scientific machines. Two good implementations are available for Control Data Corporation (CDC) hardware; Lehigh Analog Simulator (LEANS) from Lehigh University and DARE-IIIB from the University of Arizona. All these programs are written in FORTRAN and should be easily adapted to other hardware systems. A somewhat more general simulation package, called SIMSCRIPT, is available as a compiler on CDC hardware. This simulation language allows categorical as well as continuous simulation variables and has been generally accepted for logistic problems only. It is a generalized simulation language that does include CSSL; its main drawback is that IBM chooses to ignore it.

A rather more specific problem of the chemistry profession has elicited a group of programs that is closely related to the CSSL series. These programs perform a CSSL type simulation but accept problem definition as a series of chemical equilibrium equations rather than a system of differential equations. These programs are often called translators since their usual first step is to translate the system of chemical equations to the equivalent set of differential equations. After that, their numerical techniques closely parallel the CSSL type program; however, their output is chemistry oriented with automatic plots of chemical concentration versus time. These programs are discussed in detail in many of the *Computer Programming for Chemists* textbooks. Two implementations are known to be available by this author, one from IBM, and the other from Michigan State University. Most university chemistry departments have their own versions of one of these programs.

Another type of chemical simulation algorithm is also available from the RAND Corporation and is based on chemical equilibrium expressed by the Gibbs free energy function. This algorithm is limited to finding equilibrium or steady state conditions; however, it is the only technique capable of solving large systems of chemical equations. It has been successful in simulating the steady state conditions in chemical systems defined by several hundred chemical equations. The algorithm can also be implemented for use by SUMT.

A type of problem that is becoming frequent in environmental research is that called biological assay. The techniques of probit analysis, normit analysis, and logistic analysis were developed almost forty years ago by biologists and have become so standard that, except for implementation on computers, they are essentially unchanged today. The computational problem can be summarized by the following equation:

$$E(P) = \int_{-\infty}^{y} f(z)\, dz$$

where: E(P) is the expected probability of a response
f(z) is the Gaussian or normal probability function

and y is a mathematical function of the dose.

Classically y=a + b * log (dose). From this equation, one can see the origin of the common terms straight line bioassay, and log-normal bioassay.

The problem of biological assay is twofold: there is no analytical solution to the integration in the above equation, and the straight line regression is a weighted regression in which the weights depend upon the unknown values of parameters a and b. These problems are solved computationally by an iterative technique. The above formulation, while useful in laboratory research, cannot generally be applied to environmental problems because it permits the response to be dependent only upon one drug or environmental variable. In all interesting environmental problems the animals or plants are exposed simultaneously to a variety of chemicals and physical conditions. There is no mathematical or statistical reason why the quantity y in the above equation should be a simple linear function of the logarithm of the dose, but classically that is the only situation of interest to biologists; the only published methodology is likewise limited. An additional feature that can be included in the mathematical formulation of the problem is the inclusion of positive and/or negative control probabilities (the probability that an animal or plant will not respond regardless of the strength of the dose, and that without any dose there will be a response). All these concepts important to environmentalists are incorporated into a computer program written by this author several years ago for the National Cancer Institute. This program permits any linear or nonlinear function of the doses of any number of chemicals to determine the value

of y. A typical example is an analysis of variance structure where the categorical variables are the physical conditions stressing the animal (plant) at the time of observation, and the regression variables representing the concentrations of several chemical pollutants. A very important feature of this example as implemented by this author's program is that it permits the estimation of interactions among the several stressing agents and conditions.

Finally it would be beneficial to review two computer programs recommended for documentation of new programs that substantially ease the task of changing an "old" FORTRAN computer program or of understanding such a program written by someone else. Several programs are available that automatically draw flow charts of a FORTRAN program (similar programs are availabile for COBOL, ALGOL, PLI, and many other languages). Some rather elegant ones are commercially available for IBM hardware, but they are relatively expensive if not used extensively. CALCOMP has a package that produces graphic quality flow charts. For CDC equipment a FORTRAN program that flow charts a FORTRAN program was produced several years ago by the CDC users' group. This program is no longer supported, but it is available from this author. The program does work well, but its major disadvantage is that the flow chart it produces has a unique format. It should be readily adaptable to any FORTRAN system that can accommodate six alphanumeric characters per word, several moderate sized arrays, and one scratch tape. (On 16 bit 2 byte per word or 32 bit 4 byte per word machines, this can be done with double precision declarations.) A major problem with all flow charting systems is that they chart program statements and not procedural steps as does a programmer when he is designing a program. Nevertheless this program is valuable for comparing the program logic to the programmer's intended flow chart, for finding unsuspected logical loops that cause bugs, and for programmers who are not willing to effectively document their own programs. A generally available but much less powerful capability of charting all logical program structure exists within many FORTRAN compilers, in the tables that appear at the end of the compilation of each routine. The tables produced by the better compilers contain a cross-reference listing of all variable names and statement numbers which can be used to determine each use of a variable and all references to a statement.

Whenever a program of any complexity is produced, a major portion of the programming effort is spent changing the initial concept of the program. This results in a "patched-up" appearance of the program, a confusing assortment of statement numbers, and many unused statement numbers or even FORTRAN statements. In large programs these characteristics may place such a burden on the programmers that progress is extremely slow. There is also a FORTRAN program available to eliminate these problems. It is called TIDY and was produced in 1966 at the Air Force Weapons Laboratory, Kirtland Air Force Base, New Mexico. It is available through the CDC users' group and has a wide variety of options that may be chosen by the user. The eight default options are:

- All statement numbers occur in increasing order
- Only statement numbers referenced are retained
- Statement number references are updated to match consecutive ordering
- FORMAT statements are collected at the end of routines
- All FORMAT and CONTINUE statements not referenced are deleted
- Blanks are inserted and deleted to improve readability
- Comments are aligned, centered and bracketed by blank cards
- All cards are sequence-labeled with unique letter-number combinations.

The output of TIDY is a new card deck, a listing of the old and new versions of the program, and a tabulation of the changes made.

The variety of programs discussed in this paper can only be described as a selection from the accumulated program library of one person. The programs have no common theme except that their availability, or their existence, should be better known among computer users. This presentation has deliberately not been comprehensive or even moderately detailed. Hopefully it is enough to make EPA computer users aware of the tremendous libraries of computer routines that can be utilized for little more than the asking.

MSP (Model Specification Program)—An Easy-to-Use Linear Statistical Analysis System

By Larry M. Male

INTRODUCTION

Most linear statistical programs limit a user to analyses of a fixed set of statistical models. Programs for one-way, two-way, k-way, and factorial analysis of variance, covariance analysis, simple and multiple regression, etc., each assume a fixed model for estimating parameters and a fixed set of hypotheses to test. A user has no opportunity to specify hypotheses which may be of most interest to him. There do exist "general linear model" programs which allow the specification of arbitrary models, but they require that for each model, the user input a matrix of coefficients the rows of which separately specify a model for each observation in the experiment. For large data sets or complicated experiments, this approach severely limits a program's applicability. The statistical analysis program described here allows a user to easily specify any number of models that he may want to analyze.

Model Specification Program (MSP) is designed to allow simple specification of the following class of models. If an observation from an experiment depends upon the levels of c categories, $i_1, \ldots, i_c$, and the values of v independent variables, $x_1, \ldots, x_v$, then the mean or expectation of any observation is given by:

$$
\begin{aligned}
&Ey\left(i_1, \ldots, i_c, x_1, \ldots, x_v\right) \\
&= A\left(a_1, a_2, \ldots\right) + B\left(b_1, b_2, \ldots\right) + \ldots \\
&+ C\left(c_1, c_2, \ldots\right) \cdot x_1 + D\left(d_1, d_2, \ldots\right) \\
&\qquad \cdot x_2 + \ldots
\end{aligned}
$$

This means that the mean of an observation is a linear function of a number of parameters where the parameters may be indexed by any combination of the c categories. In addition, the model must be identifiable, i.e., for fixed values of the expectations, Ey, there must exist unique values for the parameters. In most cases (explained later), a reasonable reparameterization of an unidentifiable model will lead to one which is identifiable.

The necessary and sufficient information needed to completely specify a model in this form is the associated categories and independent variable number for each additive term, i.e.:

Independent Variable	Categories
0	$a_1, a_2, \ldots$
0	$b_1, b_2, \ldots$
.	.
.	.
.	.
1	$c_1, c_2, \ldots$
2	$d_1, d_2, \ldots$
.	.
.	.
.	.

MSP contains an algorithm which translates this information into an appropriate design matrix. This algorithm is constructed so that missing observations or missing cells in an experimental design present no problem so long as the resulting model remains full rank.

Since most of the concepts involved in using and programming MSP can best be understood by the generalization of an example rather than by resorting to a complicated notation, the following example is supplied. It has a factual basis in the air pollution work of NERL. Many plants produce a compound called ethelene when subjected to stress. The amount produced is a function of temperature, and the probability distribution of ethelene concentration can be shown to be approximately lognormal. A reasonable model for an observation z of ethelene concentration at temperature t is:

$$z \sim \text{Lognormal}\left(\alpha \exp\left\{\frac{-\beta}{t+273}\right\}, \sigma\right).$$

The natural log of z thus has the normal distribution:

$$y = \ln z \sim N\left(\mu + \gamma X, \sigma\right)$$

where

$$\mu = \ln \alpha$$

$$\gamma = -\beta$$

$$X = 1/(t+273).$$

In an experiment three varieties of radish were used and for each variety, two levels of stress were applied (control versus innoculation with gall disease). At each of these factors x level combinations, measurements of ethelene concentration were made for a gradient of six temperatures. The same range of temperature was used for each factor x level combination, but the actual temperatures were not identical. Six observations (plants) were measured at each temperature x factor x level combination. The fundamental model for the mean log response may be written as

Model 1

$$Ey(i_1, i_2, i_3, X) = \mu(i_1, i_2) + \gamma(i_1, i_2)X$$

where

$i_1 = 1, 2, 3$ radish variety

$i_2 = 1, 2$ stress treatment

$i_3 = 1, \ldots, 6$ replicate.

The $\gamma(i_1, i_2)$ parameters are related to the activation energy of a specific chemical reaction while the $\mu(i_1, i_2)$ parameters are related to the asymptotic production rate of ethelene at high temperature.

Examples of alternative models that an experimenter might want to examine are given in Table I.

Table I
Examples of Alternative Models

	Analytic Description	Interpretation
Model 2	$Ey = \mu$	Response does not depend upon X or any factor x level combination.
Model 3	$Ey = \mu(i_1,i_2) + \gamma(i_1)X$	Activation energy does not depend upon stress treatment.
Model 4	$Ey = \mu(i_1,i_2) + \gamma(i_2)X$	Activation energy does not depend upon radish variety.
Model 5	$Ey = \mu(i_1,i_2) + \gamma X$	Activation energy does not depend upon variety or stress treatment.
Model 6	$Ey = \mu(i_1,i_2)$	Activation energy is zero for all factor x level combinations.
Model 7	$Ey = \mu(i_1,i_2) + \gamma(i_1)X + \rho(i_2)X$, $\rho(1) = 0$	Difference in activation energy between varieties is invarient to stress treatment.

Note that if the restriction $\rho(1) = 0$ had not been made, then MODEL 7 in Table I would have been unidentifiable. This can be seen as follows. Let

$$\alpha(i_1, i_2) = \gamma(i_1) + \rho(i_2), \text{ then}$$

$$\alpha(i_1, 2) = \alpha(i_1, 1) + \alpha(i_1, 2) - \alpha(i_1, 1)$$

$$= \alpha(i_1, 1) + (\rho(2) - \rho(1)).$$

Thus all values of $\rho(1)$ and $\rho(2)$ for which $(\rho(2) - \rho(1))$ is a constant will yield the same model. If we take $\rho(1) = 0$ then $\rho(2)$ may be interpreted as the common quantity you must add to the control for each radish variety to get the activation energy for the stress treatment. In general if parameters of the form $\alpha(i_1, i_2, i_3, i_4, \ldots)$ appear in the model then MSP allows a user to specify restrictions of the form $\alpha(i_1, 1, i_3, i_4, \ldots) = 0$.

Table 1
Continued

Analytic Description		Interpretation
Model 8	$Ey = \mu(i_1) + \gamma(i_1, i_2)X$	Asymptotic response does not depend upon stress treatment.
Model 9	$Ey = \mu(i_2) + \gamma(i_1, i_2)X$	Asymptotic response does not depend upon variety.
Model 10	$Ey = \mu + \gamma(i_1, i_2)X$	Asymptotic response does not depend upon variety or stress treatment.
Model 11	$Ey = \mu(i_1) + \rho(i_2) + \gamma(i_1, i_2)X$ $\rho(1) = 0$	Differences in asymptotic response between varieties is invarient to stress treatment.
Model 12	$Ey = \mu(i_1) + \gamma(i_1)X$	Activation energy and asymptotic response do not depend upon stress treatment.
Model 13	$Ey = \mu(i_2) + \gamma(i_2)X$	Activation energy and asymptotic response do not depend upon variety.
Model 14	$Ey = \mu + \gamma X$	Activation energy and asymptotic response do not depend upon any factor x level combination
Model 15	$Ey = \mu(i_1) + \rho(i_2)3\ \gamma X$ $\rho(1) = 0$	Assuming activation energy is constant for all factor x level combinations are the differences in asymptotic response for variety invarient to stress treatment

In addition to these explicitly stated models, many hypotheses may be tested by the comparison of various models. The statistical comparison of the two models

$$Ey = \mu(i_1) + \gamma(i_1, i_2)X$$

and

$$Ey = \mu + \gamma(i_1, i_2)X$$

provides a test of the hypothesis, "the average of asymptotic response over varieties does not vary with stress treatment."

The next section provides some of the basic theory of statistical estimation and hypothesis testing which allow the construction of a general linear statistical system.

BASIC THEORY OF LINEAR STATISTICAL INFERENCE

Let $\underset{\sim}{y}$ represent a vector of the observations y $(i_1, \ldots, i_c, x_1, \ldots, x_v)$ arranged in an arbitrary order and let $\underset{\sim}{\beta}$ represent a vector of the parameters for a model. Since the expectation of each observation is modeled as a linear function of the parameters we may write:

$$E\underset{\sim}{y} = Z\underset{\sim}{\beta}$$

where the nth row of Z specifies the model for the nth observation. The matrix Z is often called the design matrix. If in our example, $Ey = \mu(i_1, i_2) + \alpha(i_1, i_2)X$, the vector of parameters are ordered as:

1	2	3	4	5	6
$\mu(1,1)$	$\mu(1,2)$	$\mu(2,1)$	$\mu(2,2)$	$\mu(3,1)$	$\mu(3,2)$

7	8	9	10	11	12
$\gamma(1,1)$	$\gamma(1,2)$	$\gamma(2,1)$	$\gamma(2,2)$	$\gamma(3,1)$	$\gamma(3,2)$

then the row of Z corresponding to y (3, 1, 4, X_n) will have a 1 in column 5 and X_n in column 11; the rest of the entries in the row would be zero.

The elements of statistical estimation and hypothesis testing are as follows. Let

$\hat{\underset{\sim}{\beta}} = (Z'Z)^{-1} Z'\underset{\sim}{y}$ = estimates of parameters

$n(\underset{\sim}{\beta})$ = number of parameters

$SS = (Z\hat{\underset{\sim}{\beta}})'(Z\hat{\underset{\sim}{\beta}})$ = sum of squares for model

$SS_o = \underset{\sim}{y}'\underset{\sim}{y}$ = total sum of squares.

Depending upon the probabilistic assumptions which can be made about the random vector $\underset{\sim}{y}$, the properties of these quantities will differ. Three cases will be considered in Table II.

Table II
Properties of Estimates Under Various Assumptions

Case	Assumptions	Properties
1	$E\underset{\sim}{y} = Z\underset{\sim}{\beta}$	The elements of $\hat{\underset{\sim}{\beta}}$ are least squares estimators of the parameters.
2	$E\underset{\sim}{y} = Z\underset{\sim}{\beta}$ $var(y) = \sigma^2 I$ (observations are uncorrelated with equal variances)	The elements of $\hat{\underset{\sim}{\beta}}$ are BLUE (best linear unbiased estimates of the parameters)
3	$\underset{\sim}{y} \sim N(Z\underset{\sim}{\beta}, \sigma^2 I)$ (y has a spherical multivariate normal distribution)	The elements of $\hat{\underset{\sim}{\beta}}$ are maximum likelihood estimates of the parameters (also other nice statistical properties)

If model 1, $E\underset{\sim}{y} = Z^{(1)}\underset{\sim}{\beta}^{(1)}$, is the fundamental model for an experiment (generally the most restrictive model which can be assumed with confidence) then

$$\hat{\sigma}^2 = \left(SS_o - SS_1\right)/\left(n(\underset{\sim}{y}) - n\left(\underset{\sim}{\beta}^{(1)}\right)\right)$$

is an estimate of σ^2. Under case 3 $(n(\underset{\sim}{y}) - n(\underset{\sim}{\beta}^{(1)}))\hat{\sigma}^2$ has a chi-square distribution with $n = (n(\underset{\sim}{y}) - n(\underset{\sim}{\beta}^{(1)})$ degrees of freedom.

An estimate of the covariance between two estimates of parameters is given by

$$\widehat{cov}\left(\hat{\beta}_i, \hat{\beta}_j\right) = \hat{\sigma}^2 (Z'Z)^{-1}_{ij}.$$

Under cases 2 and 3, $\underset{\sim}{a}'\hat{\underset{\sim}{\beta}}$ is the best estimate of $\underset{\sim}{a}'\underset{\sim}{\beta}$ (a linear combination of the parameters) and its variance can be estimated by

$$\widehat{var}(\underset{\sim}{a}'\hat{\underset{\sim}{\beta}}) = \hat{\sigma}^2 \underset{\sim}{a}'(Z'Z)^{-1}\underset{\sim}{a}.$$

Under case 3, a $(1-\alpha)\times 100\%$ confidence interval for $\underset{\sim}{a}'\underset{\sim}{\beta}$ is given by

$$\underset{\sim}{a}'\hat{\underset{\sim}{\beta}} \pm t_{m,\alpha} \cdot \sqrt{\widehat{var}(\underset{\sim}{a}'\hat{\underset{\sim}{\beta}})}\ ,$$

where $t_{m,\alpha}$ is the $(1-\alpha)^{th}$ percentage point of the students t distribution with m degrees of freedom. The program MSP calculates the quantities $\hat{\underset{\sim}{\beta}}$, $n(\underset{\sim}{\beta})$, SS, and $(Z'Z)^{-1}$ for each model specified.

Hypothesis testing is essentially the statistical comparison of two models. If

Model 2 $\quad E\underset{\sim}{y} = Z^{(2)}\underset{\sim}{\beta}^{(2)}$

and

Model 3 $\quad E\underset{\sim}{y} = Z^{(3)}\underset{\sim}{\beta}^{(3)}$

are two models such that model 3 is a submodel of model 2 (i.e., the subspace spanned by the columns of $Z^{(3)}$ is a subset of the subspace spanned by the columns of $Z^{(2)}$), then under case 3 the quantity $(SS_2 - SS_3)/\sigma^2$ has a noncentral chi-square distribution with $n(\underset{\sim}{\beta}^{(2)}) - n(\underset{\sim}{\beta}^{(3)})$ degrees of freedom. If model 2 is a submodel of model 1 (the fundamental model), as it should be, then the statistic

$$F = \frac{(SS_2 - SS_3)/\left(n\left(\underset{\sim}{\beta}^{(2)}\right) - n\left(\underset{\sim}{\beta}^{(3)}\right)\right)}{\hat{\sigma}^2}$$

has a central F distribution when model 3 is true and a noncentral F distribution when model 3 is false but model 2 is true. The F statistic thus provides a test of the hypothesis

H_o : model 3 is true

H_A : model 2 is true

The test procedure is given by:

decide in favor of model 2 if

$$F > F_\alpha \left(n\left(\underset{\sim}{\beta}^{(2)}\right) - n\left(\underset{\sim}{\beta}^{(3)}\right), m \right)$$

where $F\alpha$ (u, v) is the $1\text{-}\alpha)^{th}$ percentage point of the central F distribution with u degrees of freedom in the numerator and v degrees of freedom in the denominator.

MSP allows the comparison of models and will perform the calculation for an F test. When a user indicates to MSP that two models should be compared then the following quantities are computed.

Degrees of freedom for hypotheses

$$= n\left(\underset{\sim}{\beta}^{(2)}\right) - n\left(\underset{\sim}{\beta}^{(3)}\right)$$

sums of squares for hypotheses

$$= SS_2 - SS_3$$

mean square for hypotheses

$$= \left(SS_2 - SS_3\right) / \left(n\left(\underset{\sim}{\beta}^{(2)}\right) - n\left(\underset{\sim}{\beta}^{(3)}\right)\right).$$

To construct an F test for an hypothesis, the user specifies the two models to be compared and the two models to be used to compute the error mean square (usually $(SS_o - SS_1)/(n(y) - n(\underset{\sim}{\beta}^{(1)})$, but for subsampling experiments this might not be the case).

MSP also employs a routine for calculating the central F distribution function. Thus for each test of hypotheses the probability of a Type I error (rejecting the hypotheses when it is true) is calculated.

ALGORITHM FOR CONSTRUCTING THE DESIGN MATRIX

The main algorithm in MSP which makes the program easy to use is the one which constructs the design matrix from the minimal input needed to specify a model. The technique used is most easily understood by the generalization of an example. Consider model 1 of the example,

$$Ey_n\left(i_1, i_2, i_3, X_n\right) = \mu\left(i_1, i_2\right) + \alpha\left(i_1, i_2\right) X_n .$$

The design matrix will have n(y) (number of observations) rows and $n(\underset{\sim}{\beta}) = 12$ (number of parameters) columns. In any row the number of nonzero entries will be equal to the number of additive terms in the model: two in this example. The nth observation will be associated with one parameter for each additive term in the model. If we have constructed a serial ordering for the parameters then the elements of Z are given by

$$Z_{nj} = 1 \quad Z_{nk} = X_n$$

$$Z_{nm} = 0 \quad \text{for } m \neq j, k$$

where j and k are the serial numbers of the parameters associated with the first and second additive terms for the nth observation.

MSP first constructs a serial ordering for the parameters and then consecutively constructs each row of the design matrix by examining the associated levels of the categories for each observation. An index vector keeps track of which parameters are not used. If any parameters are not used, then the corresponding columns are deleted from the design matrix. Thus if there are missing observations or missing cells in an experimental design only those parameters which can be estimated from the available data will be included. If parameters like α (i_1, 1, i_3, i_4) occur in the model with a restriction such as α (i_1, 1, i_3, i_4) = 0, then this information is stored in the index vector and the appropriate columns are dropped from the design matrix.

In actual practice MSP does not store the design matrix but instead accumulates the appropriate cross

products from the matrix $Z'Z$ and the vector $Z'y$ as each row of the design matrix is constructed. Since many of the elements of a row in the design matrix may be zero, only those elements of $Z'Z$ and $Z'\underset{\sim}{y}$ which will change are accumulated. Referring to the example, only the following replacement statements are executed:

$$(Z'Z)_{im} = (Z'Z)_{im} + Z_{ni} \cdot Z_{nm}$$

$$\text{for}_{im} = jj,\ jk,\ kk$$

$$(Z'\underset{\sim}{y})_i = (Z'\underset{\sim}{y})_i + Z_{ni} \cdot Y_n$$

$$\text{for } i = j,\ k$$

Since $Z'Z$ is a symmetric matrix only the upper diagonal is stored.

If an experiment contains several dependent variables for which the same analysis is desired, then it is a waste of time to construct the same design matrix for each separately. For this reason MSP calculates the parameter estimates and sums of squares for each dependent variable using only one construction of the design matrix.

SUMMARY OF FEATURES OF MSP

1. MSP handles experiments defined by any number of categories, any number of independent variables, and any number of dependent variables. The maximum size or complexity of an experiment is limited only by the maximum amount of core storage available.

2. Required input to the program is:

   - Raw data and user supplied format specification.

   - Number of observations, categories, independent variables and dependent variables.

   - Model specification cards. Each model requires one card for each additive term. On each of these cards the user specifies the independent variable number, the category numbers which index the parameters, and information on restrictions of the form $\alpha\,(i_1, i_2, i_3, i_4, i_5, i_6, \ldots) = 0$.

   - Model comparison cards. The models are serially numbered so that a user merely specifies a pair of integers for each model comparison card.

   - F-test cards. The user specifies two pairs of model numbers: one for the numerator and one for the denominator of the F statistic.

3. Output from the program is

   - For each model specified an analytic expression of the model is written on the output. For example:

     EY = A(I1, I3) + B(I2) · X1
     + C(I1) · X2.

   - For each model the estimates of the parameters for each dependent variable are printed in the form given below:

#	Symbol	Values			
1	A(1,1)	XXX	XXX	XXX	XXX
2	A(2,1)	XXX	XXX	XXX	XXX
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
	A(I1, I3)	XXX	XXX	XXX	XXX
	B(1)	XXX	XXX	XXX	XXX
	.	.	.	.	.
	.	.	.	.	.
	.	.	.	.	.
	B(I2)	XXX	XXX	XXX	XXX
	.	.	.	.	.
	.	.	.	.	.
	.	.	.	.	.
n($\underset{\sim}{\beta}$)	C(I1)	XXX	XXX	XXX	XXX

   - For each model the number of parameters and the sums of squares for the model are printed.

   - If the user asks for the $(Z'Z)^{-1}$ matrix, it is printed in the form:

     COV (1,1) = XXX
     COV (1,2) = XXX
     .
     .
     .
     COV (n($\hat{\underset{\sim}{\beta}}$), n($\hat{\underset{\sim}{\beta}}$)) = XXX

Multiplication of these values by the users selected estimate of σ^2 yields the variances and covariances of the parameters.

- Model comparison table which presents the degrees of freedom, sum of squares, and mean square for each hypothesis.

- Tables which present the results of the F-tests. Included are numerator and denominator degrees of freedom, F statistic, and probability of a Type I error.

4. Techniques such as normalizing the $Z'Z$ matrix before inverting are used to minimize rounding error.

5. The programming language is FORTRAN IV for the Oregon State University CDC 3300.

Additional features such as automatic confidence intervals for parameters, three decision-decision procedures, and even more simplified input are contemplated for the near future. The program should be completed along with a user's manual by the end of 1975.

CONCEPTS VERIFICATION TESTING OF THE ENVIR SYSTEM

By John W. Scotton

PLANNING AND COORDINATION

The contract with Gulf Universities Research Consortium (GURC) involved joint Environmental Protection Agency/Office of Research and Development (EPA/ORD) and GURC planning and coordination. It also involved the selection of three categories of EPA-generated data considered representative of, and suitable for, demonstration/evaluation. The following categories were selected in agreement with the program participants, and a firm plan for delivery of data tapes to GURC was developed. At the same time, priorities were established and continuing coordination arranged. By personal visit to the GURC facility and through active participation in the creation of a data base out of Data Category 1 and thereafter personal query/response of the first "Dynamic File" on UNIVAC 1108 at the University of Houston, a firm project plan was implemented.

Presentations, discussions, work sessions and demonstrations were organized and coordinated to allow maximum visibility and exposure to EPA divisions; emphasis was placed upon interagency liaison. Demonstrations were scheduled at Waterside Mall and at Research Triangle Park. The need for evaluation of those demonstrated applications to the Agency required followup discussion, debriefing and continuing consultation with the GURC scientists.

These demonstrations/evaluations were, in this pilot program, limited to application of the Environmental Information Retrieval (ENVIR) system. This ENVIR system is the primary module in the Environment Dependent Management, Process Automation and Simulation (EDMPAS) system. As a consequence of extensive briefings, discussions and training sessions, EPA/ORD staff have acquired sufficient in-depth knowledge of the total systems concept to provide evaluation comment.

DATA CATEGORY 1

National Soils Monitoring Program Data

Mr. Elgin Fry, Office of Pesticides Monitoring, provided a magnetic tape to GURC containing the National Soils Monitoring Program Data for the year 1971. This data tape consisted of data related to place, time, crops, pesticides, method of application, amounts, and so forth to a total of 117 fields. The existing transaction-oriented type printout was also provided.

This data tape was reformatted by GURC in accordance with the methodology reported and, in addition, the fields were entirely restructured so that pesticides became the operative element in each record rather than the location at which the pesticide was monitored. This restructuring of the file rendered the data more useful insofar as the new structure coincides with the primary monitoring requirement. This results in ENVIR records containing 25 fields. While providing all of the original information, such records could have been punched on almost half the original number of cards. Data entry, involving staff and materials, is considered important in terms of system cost effectiveness.

The restructuring of this file, effectively accomplished through simple interface software, permits an unlimited number of pesticides and crops to be incorporated in the data base. The existing restriction of 15 pesticides and 3 crops, limitations imposed by the current system, need no longer be tolerated.

The GURC team produced computer printout exactly duplicating the original information contained in the conventional output. Additional queries, pertinent to pesticides monitoring, were demonstrated. All queries requested by EPA staff through batch-mode terminal were scrutinized and found totally responsive.

In brief, this demonstration indicated that:

- Data entry is more efficient
- Query capability is unlimited
- Hierarchical printout makes the information more meaningful
- Information produced through certain queries had not been available before.

A typical comment made by one staff member was that: "The information produced from his query,

costing 23 cents, would have taken two days of staff time for response."

The complete file, dictionary structure, dimension statements and typical queries are illustrated in a bound volume of computer output with a report on the methodologies applied and the system User Manual. The computer output was duplicated on UNIVAC 1100 series and IBM 370 series computers.

DATA CATEGORY 2

Federal Water Pollution Control Surveillance Data

Mr. Lee J. Manning, Office of Water Quality, provided a magnetic tape and printout of Water Quality Data (STORET) containing Potomac River Basin data. For purposes of this pilot program demonstration, it was considered more cost effective to repunch rather than reformat the tape. From the cards, punched in free field ENVIR format, a new tape was built. The Dynamically Structured Data Information File relevant thereto was created on UNIVAC 1108 at the University of Houston for purposes of checkout and thereafter recreated on IBM 370/165 at Bethesda, Maryland through batch-mode terminal.

This methodology demonstrated the ENVIR skip and dupe capability to reduce data entry. The ENVIR facility for repetitive data entry (correction statements) was also applied. The complete Boolean algebraic capability permits the partitioning of data, thus allowing the application of a correction statement specific to a set or subset of data.

GURC staff considers that a record should consist of tests of water characteristics at a particular TIME and PLACE. This permits partitioning of the Data Base with a single Boolean Expression so that all localities where a pollutant level is in violation, even for only part of the year or month, may be extracted.

In addition to the demonstration, a work session discussion was held with Mr. Sam Conger. It is evident that, if interfaced with STORET, ENVIR offers capabilities to expand the query and output facilities.

The data base, dictionary structure, dimensions, typical query responses and appropriate reports for this Data Category 2 reside with the writer.

DATA CATEGORY 3

Emission Inventory Point Source Data

Mr. Gerald Nehls, Office of Air Quality Planning and Standards at Research Triangle Park (RTP), North Carolina, provided information related to existing data banks at RTP (e.g., SAROAD, Emission Point Source and Area Source) and allowed GURC to select a data tape. GURC responded with a request for Point Source data, data pertinent to the EPA mission, representative of a broad geographical area, and a reasonable expanse of time. RTP therefore dispatched a tape covering various time frames and containing representative data from Hawaii to Maine and Florida to Alaska.

The computer programs presently used at RTP for Point Source data retrieval appeared to consist of six separate records for each Point Source. GURC therefore requested that the tape contain the data such that each set of six records could be considered as one. RTP complied. As a consequence, this reformatted Point Source data provided a logical structure permitting the full capabilities of ENVIR to be demonstrated. Prepared front-end software was employed to convert the tape, IBM/EBCDIC code to UNIVAC FORTRAN readable. At the same time, an additional routine to convert fixed field to ENVIR readable was applied.

In this same UNIVAC Run Stream, an ENVIR Dynamic File was built containing 85 fields (descriptors). Fifty field were automatically excluded from this Dynamic File. Forty-eight of these fifty fields were excluded because they were essentially redundant or private to the existing system (e.g., card numbers, actions, etc.). ENVIR has no need for such fields. They could, nevertheless, be accommodated in ENVIR but they add nothing to the query capability. The remaining two fields were extracted and a separate file created simply to demonstrate that, when necessary, the computer can be used to provide a "book."

This main ENVIR Dynamic File of scientific data contains all of the original spatial, temporal and pollution data in the original six file cards. The Dynamic File created on UNIVAC 1106 at Mississippi State University was read directly into the UNIVAC 1110 at RTP for purposes of demonstration. By sample query and response, it was demonstrated that the system can be used to predict ambient air quality design sampling

standards, control strategies, pollution violations, stack characteristics of regions, and so forth. Subsequently, this same Dynamic File was queried through batch-mode terminal at the Waterside Mall line connected to the UNIVAC 1110 at RTP.

The data base, the dictionary structure, discussions and typical query responses, and appropriate reports were delivered to EPA/OKD for this Data Category 3.

DEMONSTRATION

On Monday, June 3, 1974, the GURC Information Staff provided a three-hour presentation of the EDMPAS system concepts to representatives of EPA at RTP. By prearrangements with GURC, complete sets of zerox copies of viewgraphs and slides were bound together with general narrative to provide a meaningful document. A presentation agenda was agreed upon and distributed prior to this visit.

Following the presentation, RTP arranged for a full demonstration of the Dynamic File created out of Point Source Emissions data. Using the standard terminal with the CRT photographed and line-connected to several standard T/V outlets through the audience in the main conference room, it was possible for any number of the audience to write or call for queries, input through teletypewriter by an EPA staff member, and to receive immediate response to query by CRT. As a consequence, it was possible to conduct a practical systems demonstration to a large audience with full audience participation.

This same Dynamic File, left resident on disc for a few days, was again queried through the batch-mode terminal in Waterside Mall at the Office of Pesticides Monitoring. In parallel Dynamic Files mounted on IBM 370/165 at Bethesda, Maryland were similarly queried. In this manner, it was possible for a representative group from several divisions within the agency to participate directly. In addition, guests from other U.S. Federal Agencies were invited.

THE ENVIR SYSTEM CHARACTERISTICS AND CAPABILITIES DEMONSTRATED AND EVALUATED

The following summary of characteristics and capabilities included within the GURC intermediate reports was carefully scrutinized and used for evaluation. For a potential user to check any or all of these requires some brief study of the User Manual. A perusal of this manual coupled with a document entitled "ENVIR, An Illustration Check Deck," will demonstrate the simplicity of application and operation; it will also demonstrate the flexibility of this ENVIR module in the creation of dynamically structured data/information files of multidisciplinary data.

To fully understand and appreciate the new concept in information management offered in the design of the EDMPAS system, of which ENVIR is the primary module, requires study of "The Systems Brochure" which describes the system in its entirety from a user standpoint.

The three data categories selected and converted to Dynamic Files (for which complete computer runs are available for inspection) fully demonstrate most of the characteristics of the system projected by GURC. The EPA/ORD evaluation was conducted in conjunction with the divisions represented in an entirely satisfactory manner. Furthermore, staff participation in this entire process permits the expression of the view that all of the following were fully demonstrated with the system:

- Accommodates and readily adapts to the storage and retrieval of data records consisting of essentially unlimited numbers of descriptors and states (ranges, resolution) of variable bit length

- Operates as a completely free field system wherein no card column requirements are imposed; i.e., data can be received and automatically integrated into the files without any predetermined fixed card column format requirements

- Accepts data which is alpha, alphanumeric, numeric or coded without restrictions

- Accepts data regardless of its managerial or scientific disciplinary content

- Dictionaries for English/binary/English translation are built automatically from input data; the user is not required to furnish a predetermined and inflexible thesaurus in advance

- Permits any data entry in the dynamic file to be corrected, any record to be deleted, and more data to be added anytime during the life of that dynamic file; new descriptors may be

added to the dynamic file anytime and combining disparate batches of data punched in different format

- Full Boolean capability is available, and all three Boolean operators AND, OR NOT may be used while retrieving data

- All the descriptors in the dynamic file are retrievable

- Permits the selective retrieval of any or all records in the dynamic file without instructional software expansion, without search and compare procedures, and without the inversion of any descriptors or descriptor-states in the file and, therefore, without associated file expansion

- Operational simplicity: permits direct access to the file by noncomputer professionals and their utilization of the data for analytical, interpretative, documentation, and management functions

- Several options to format the query output in the desired hierarchically indented and classified manner

- Fast binary sort of the selectively retrieved data

- The user has complete control of designing the structure of the dynamic file to best suit his needs; he decides what descriptors are to be incorporated into the dynamic file, and how each descriptor is to be handled

- Simple procedures to allocate core to the various arrays used by ENVIR thereby adjust the program dimensions to fit the particular data bank and optimize program performance

- In the ENVIR compressed file, compression of the data is achieved by storing data in binary form and structuring the files such that every bit in core assigned to data storage is available for that purpose

- Operation speed: the fast query response was demonstrated during the demonstrations on EPA hardware.

It should be emphasized that ENVIR does not use the conventional binary codes in the Dynamic File. Rather, each unique field state is assigned an ordinal binary number representation. Thus, a data field with two distinct states uses only one bit in the form of 0 or 1. Consequently, a geometrical progression is created so that 4 unique states uses 2 bits, 8 states, 3 bits, 16 states, 4 bits, ... a million + states, 20 bits. Each distinctive state is assigned a code in this fashion whether it is 1 or 126 characters long. Each distinctive alpha character state is stored once in a conventional binary coded file. Numerical states are assigned only ENVIR (BIT-VECTOR) codes.

The system is written in FORTRAN V and is therefore essentially machine-independent. Note the availability of both IBM and UNIVAC versions of this ENVIR system. Files created on UNIVAC 1108 were directly transferred to UNIVAC 1110. The dictionaries are separated on disc from the compressed data files and thus offer computer system independence. It is possible, by front-end software, to translate the dictionary while transferring the compressed file directly from machine to machine. While these system facilities were not checked as part of this evaluation, it can be said that this ENVIR module of the EDMPAS system appears to offer:

- Machine independence

- Computer system independence

- Interlinguistic possibilities through use of correction statements.

THE EDMPAS SYSTEM

The total EDMPAS system is described in a brochure available for issue. To conduct an evaluation of ENVIR, the primary module in this EDMPAS system, it was considered necessary to acquire a full understanding of the total systems concept. It will be noted that as a stand alone system, ENVIR has the capability to create dynamically structured data information files to the user's specifications and for file content addressability through unrestricted Boolean query. This permits system application in a cybernetic mode.

To take full advantage of this research mode, the ENVIR system at Mississippi State University is directly interfaced to display modules and processing modules of the EDMPAS system. The User Manual demonstrates the

commands permitting direct input to modules. For purposes of continuity in this report, a summary of EDMPAS and a system diagram are included (Figures 1 and 2).

Those presentations provided by GURC and subsequent work sessions and consultation coupled with participation in file creation at the University of Houston have provided EPA/ORD with all necessary background information appropriate to a general evaluation.

A study of the documentation of the system and an examination of the reports based upon system applications already undertaken and completed clearly demonstrate the applicability of this system to agency needs. The EDMPAS system offers a capability to the agency whereby multiinstitutional, multiagency, multidisciplinary data can be combined to produce data bases, which can be correlatively analyzed by scientists and the resulting information synthesis conducted by EPA scientists in conjunction with agency and institutional scientists external to EPA. In this way, scientific credibility and public credibility of management decisions can be projected and demonstrated.

THE EDMPAS SYSTEM: HOW TO ACHIEVE AN IN-DEPTH SYSTEMS ANALYSIS

The complete system as previously described is mounted and operational at the UNIVAC Computing Center at Mississippi State University at Starkville, Mississippi. It should also be noted that the primary module ENVIR is mounted at the University of Houston UNIVAC 1108 and the Rice University IBM 370 system. During these demonstration evaluations, dynamic files were created on UNIVAC and directly read into the RTP UNIVAC 1110.

To conduct an in-depth analysis of the EDMPAS system requires only that:

- Dynamic Files of environmental sciences data be created out of systems data tapes.
- Such dynamic files contain multidisciplinary data with a large number of fields and thousands of records with spatial and temporal coordinates included.
- The data should include descriptors and states appropriate to (1) objective analysis, (2) graph clustering analysis, and (3) Charanal processing, etc.
- These files on tape can then be transported to Mississippi State University where they can be read into the EDMPAS system for in-depth analysis by the GURC staff in conjunction with the Agency scientists.

CONCLUSION

In line with the GURC objectives, this evaluation of ENVIR conducted by EPA/ORD was designed to establish the applicability of the ENVIR system to the collation, sorting, correlative analysis and interpretation of selected categories of EPA data. As a consequence of successful demonstration/evaluation, it is considered appropriate to recommend in-depth systems analysis and application of the EDMPAS system.

GURC has developed this information management system to that degree where, in conjunction with a user agency, a systems design can be jointly determined for immediate application to Agency management needs.

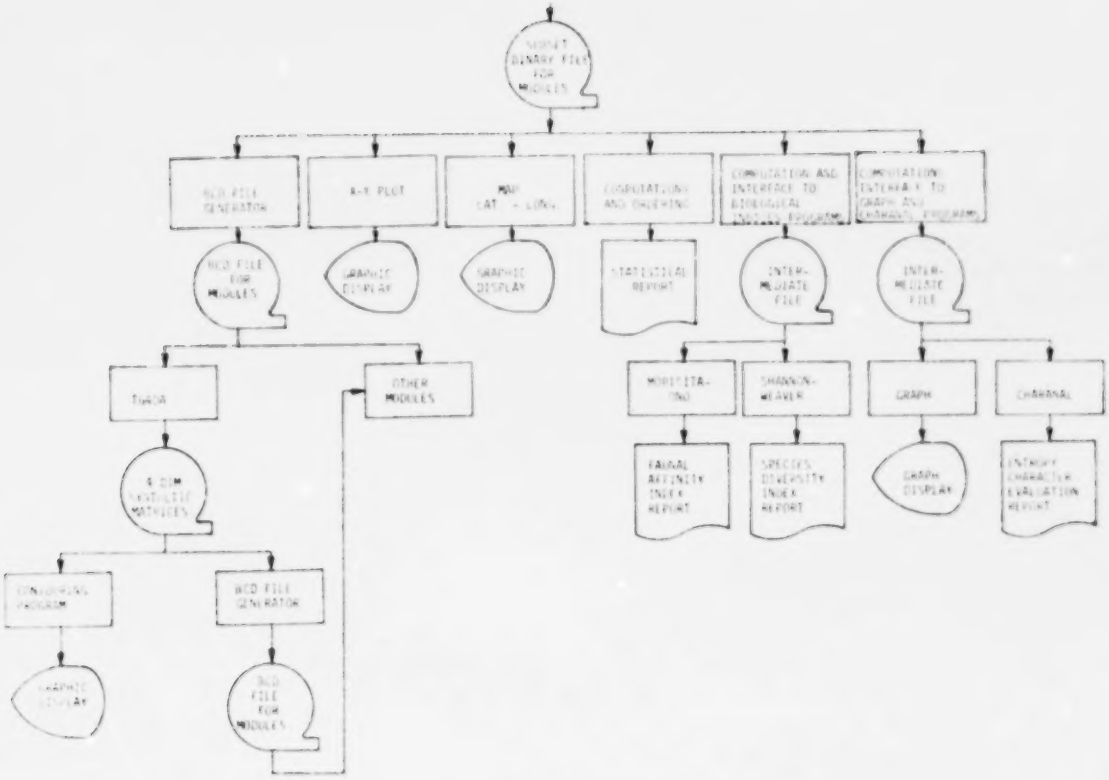


Figure 1
A Summary of Environment-Dependent Management Process Automation and Simulation (EDMPAS)

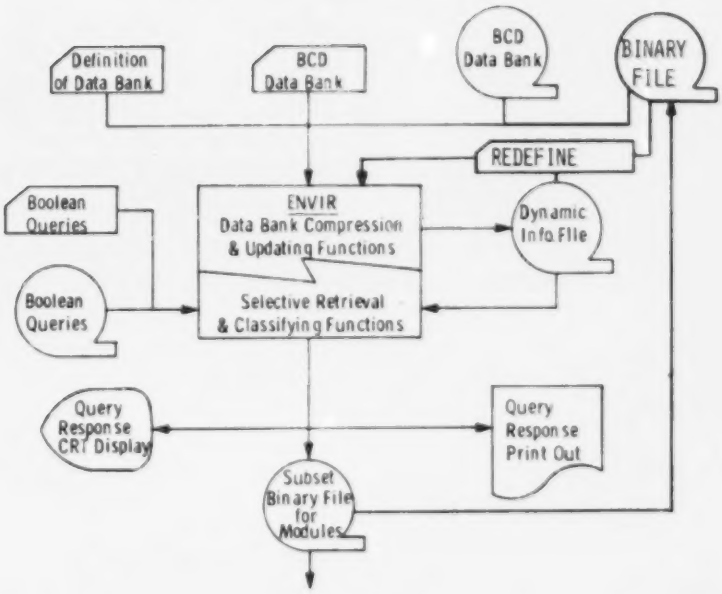


Figure 2
EDMPAS System Development Status

Reproduced from best available copy.

SEAS - THE STRATEGIC ENVIRONMENTAL ASSESSMENT SYSTEM

By Edward R. Williams

INTRODUCTION

The system to be discussed is the Strategic Environmental Assessment System (SEAS). SEAS is a comprehensive analysis and impact forecast system (see Figure 1). As such, it can serve as a coarse-grain, early warning system that will allow the tracing of impacts of policy actions in economic, energy and environmental areas. Design goals of the system include:

- Projection of demographic and economic variations to derive impacts on pollutant residual levels
- Estimation of treatment costs for major end-cost sectors
- Prediction of secondary effects and reactions due to various environmental quality levels.

For the set of goals outlined in Figure 2, the expected characteristics of the system are detailed as:

- National forecasts for a period of 10 to 20 years
- Forecasts disaggregated to regions, such as states
- Forecasts that include economic levels with different abatement levels and timing assumptions
- Comprehensive forecasts to include industrial, consumption, transportation and disposal sources in order to be a comprehensive state of emission report
- Computer-based, allowing rapid parametric or alternative assumption assessments.

Figure 3 provides the original concept of the SEAS system as designed in the fall of 1972 and serves as a basic model for discussing present as well as future development plans. The highest line on the general form of the model shown in Figure 3 reveals a diverse and large data base and a significant interpretive capability (shown as expert opinion). Most of the data changes to develop an alternate set of scenarios require analyses for engineering expectations plus feasibility and timing checks.

The major line of interest is in the center of Figure 3. From left to right the concept calls for:

- The inputs, or change agents, for a particular scenario. These are factors that would impact the envirnoment and its management. They include such data as trends in population, industrial growth and shifts, energy conservation measures, and technologic changes in goods production or abatement procedures.
- The processes shown in Figure 3 are the activities of the socioeconomic system that respond to the change agents. These processes include extraction, production, distribution, consumption, and disposal. Thus the full set of man-implaced processes is represented for several years' forecast levels, providing both a measure of the economic supply and demand plus a basis for projecting gross environmental pollutants.
- SEAS produces annual emission levels for the the set of environmental residuals shown in Figure 3. These residuals are provided by industry and consumer activity at several levels of interest including gross or untreated residuals, net residuals (left after treatment), secondarily produced residuals, and recycled materials. About 125 types of pollutants are represented in the general areas of air, water, solid waste, pesticides, and radiation residuals.
- From the residuals output there are several effects: costs associated with abatement procedures, ambient levels, and other benefits. This concept area and the following one are concept areas of SEAS that remain only partially developed at the present time.
- In addition, there are reactions due to exposure to residuals or other anticipated impacts. This area has had little development to date

although some analysis of regional economic distribution change is now under study.

- Finally, of course, SEAS would generate many reports with each element to include adjustment feedbacks to other elements. This completes the conceptual procedure.

DEVELOPMENT HISTORY

From the fall of 1972 until the spring of 1973, the study group assembled a test procedure, using submodules where available, discovering where design objectives were unfeasible or the proposed product still under development, and generally rating the problem areas. A working computer system for testing ideas and a plan for a reduced-scope operational model were developed; a set of surveys of available theory and supporting data was produced. From these results, a set of plans detailing procedures for the early development of a product to be used in EPA analyses was completed.

The results of the first phase were constructive enough to allow an operational model, called the Prototype, to be tested in January 1974 and with refinement, this Prototype SEAS is now being used for several studies in EPA and other agencies. Much of this article is based on the Prototype SEAS; however, note that development is continuing to expand the scope of the forecast tool as well as refine its present components. The Prototype of March 1974 has been updated to include three new modules; and by next summer, the third phase of development will be completed to include all improvements planned at this time. It should be emphasized that Phase III is not only a time for development but also is the first major phase of user applications. The results of these analyses will probably direct the areas of further SEAS development after next summer.

THE PROTOTYPE SYSTEM

The following discussion briefly describes the Prototype and its parts. Figure 4 shows that the system is complex and is designed to have a great deal of interaction and feedbacks among its program modules. Its major areas consist of the following:

- Economic input/output models
- Energy budgets
- Envirnomental residuals
- Abatement costs
- Consumption residuals
- Spatial assignment and so forth.

By judicious treatment of scenario input data, the previous areas can be assessed. However, any model of this size also requires a good deal of iteration and adjustment to fully balance effects; thus, the operating analyst must expect to spend a fair amount of time in analysis of inputs to properly structure those change agents that were called for in the concept.

Figure 5 is a pictorial representation of the SEAS Prototype system. The first major module takes macrodemographic and macroeconomic inputs and, using an input/output procedure, develops a national economic forecast for each year to 1985. The economic model called INFORUM was developed by Clopper Almon at the University of Maryland and represents the national economy as 185 product sectors. Even this level of detail is not sufficient to fully represent those industrial processes that are needed to show environmental and energy-related effects; therefore, further detail is added to give an additional 100 production subcategories. This part of SEAS provides dollar volume outputs that can be associated with physical noneconomic outputs for the industrial segment of national activities.

The next block in Figure 5, RESGEN, provides the set of environmental residuals associated with industry. The first purpose of this module is to determine the gross residuals that are produced by the economic process. Next, it associates those environmental abatement processes with the gross residuals to modify the levels and forms of pollutants that are released to the carrier media. An obvious need is to represent information, such as time phasing of emplacing processes, to meet environmental regulations such as BPT and BAT. Notice that transformations can be intermedia; that is, treatment of an air residual could give both water and solid waste residuals.

Two bookkeeping routines are part of SEAS. Energy demands are associated with the economic output in order to develop an energy forecast by nine major fuel types for each year. The second bookkeeping module uses the economic level and the degree of environmental treatment in each sector to calculate capital costs and annual O&M costs due to the level of abatement.

All of the data generated thus far are at the national level; as an option, the SEAS Prototype can regionalize all industry data to federal regions and to

states. The major mechanisms used are base year data from county-city data bases and the Department of Commerce OBERS regionalization shift-share forecast system. While these are the state-of-the-art forecast procedures, overrides can be introduced where better data are known.

Turning from the industrial drivers of SEAS, three other residual producing procedures are derived at state levels and these are summed to provide national totals. These are transportation, space and water conditioning, and consumer solid waste disposal. In each case a set of residuals based on alternative abatement strategies is forecast and costs of abatement, or of disposal, are developed.

This description completes the system overview except for two modules. The first is a macroforecast of land use acreage based on the industrial production levels. The second is the report generator that summarizes all of the data of all assessment modules. These data are in terms of national and regional totals.

The environmental residuals generator for industrial processes is presented in Figure 6. RESGEN is driven by two files, both representing output levels on an annual basis from the economic input/output models. For a given base year, the system is loaded with a set of multipliers that projects the gross level of residuals that would be produced at the given production level according to the environmental state of the art then in place. A set of other coefficients is provided to reflect the environmental abatement processes for that base year. For example, for a water effluent the industry treatment can be divided into the fractions that represent fully internal (onsite) treatment, pretreatment onsite followed by municipal plant treatment, and no treatment process. Associated with each of the treatment options for 1971 is a coefficient that represents the amount that is captured and transformed, leaving the amount of residual that escapes in its original form to media. Additionally, the materials that are captured are subdivided into two groups: material that is recycled to the economic process and material that is transformed into different, or secondary, environmental residuals.

An examination of the right-hand column of Figure 6 reveals additional data elements considered in future years. First, there may be economic technology changes. Such changes would modify the gross emission coefficient; an example would be more efficient use of mercury in a manufacturing process resulting in a reduction of the level that is vaporized as part of the economic production system. Additionally, any of the abatement factors may be modified over time through such means as treatment, improved efficiencies of amount of residual captured, increased recycling, and so forth. This is obviously where the data are placed to reflect meeting goals of BPT, BAT and so on. The system is flexible enough to allow a full representation of time series changes and to allow easy modification of the years when requirements are to be met. With the introduction of these changes in the residual coefficients, the calculation procedures are much the same as the base years.

Figure 7 is a list of the primary residuals; the number of secondaries is slightly less than this list. Of course, many of the industrial sectors produce similar residuals so that the number of residual coefficients is on the order of thousands from industry sources alone.

SEAS COMPUTER CHARACTERISTICS

Figure 8 is helpful in discussing the SEAS total system from a computer operations sense. In Figure 1, SEAS was represented as a set of independent modules connected by shared data bases. Figure 8 contains the list of programs in order of operation. Of the 17 programs, the main constraints are contained in UMINFOR, the input/output program, and in DISAG, the assignment of economic residuals to states. UMINFOR requires the largest core, which is solved in Phase III by overlaying the program down to about 175K bytes. DISAG has the longest running time and is the only program requiring data from tapes. The 15-minute running time required now is for regionalized data at state level. We are planning to go to AQCR's and SMSA's by next June; this detail could triple the running time. Fortunately, the need to regionalize occurs in only about one-tenth of the analyses.

Each program can be operated on call, but other options have been set up for users including one called SUPERSTREAM. It does everything except DISAG, using 850 lines of JCL.

SEAS is in FORTRAN IV and resides on private disc packs on the Optimum Systems, Inc. 370/158

computer. It has several users, including two government agencies outside of EPA.

SEAS PLAN III

Figure 9 shows SEAS Phase III scheduled to be operating in June 1975. It follows the same patterns as the present system but also includes:

- An improved macroeconomic driver
- Expanded regionalization to SMSA, AQCR, and river basins
- Expanded cost modules
- Expanded energy and scarce materials budgets
- A first set of damage functions
- An improved management system to better automate feedbacks and internal data assistance.

EXAMPLE OUTPUTS

SEAS is set up with a default data base representing the historical projection data base overlapped with known legislative effects. Some of the major assumptions are presented in Figure 10. Using this set of change agents, a set of primary pollutant emission levels predicted by SEAS is depicted in Figure 11. The first numerical column gives the absolute value for the base year 1971. The next three columns give annual growth rates for the specified three-year periods. There are quite different patterns for the time periods representing the timing and input of the various legislation actions. Most of the air emissions and all of the water pollution levels drop in effectiveness from 1977 to 1980, reflecting the input assumptions that the first level treatment facilities are in place by 1974 and that the second level facilities are negligible prior to 1980.

A more graphic presentation of these effects is presented in Figure 12 through a set of graphs. The solid line is the data for base cost extended to 1985. The broken line indicates the result if industrial process abatement effectiveness and transportation controls are held at the effectiveness levels and the installation incidence given for the base year 1971. Notice in the first three graphs concerning air emissions that there is a definite inflection in the solid line where the industrial abatement processes reach full effect in 1977. The first two emissions are primarily from industrial sources, and from 1977 to 1985, they are directly in proportion to the projected industrial growth. The line representing carbon monoxide is much smoother since the full effect is not realized until older automobiles are phased out. Note that for the first few years the dashed line drops due to abatement techniques begun in 1968.

Figure 13 provides a set of 13 energy conservation measures and the expected static savings of energy in Btu's. Additionally a change in electrical power generation toward greater use of coal is assumed. Totaling the expected savings provides a fair representation of the energy conservation half-and-half plan. This set of measures is considered speculative and on the outer bonds of economically feasible levels.

Figure 14 details some of the impacts expected from energy conservation measures. The value for each scenario and each statistic in 1971 is set at 100. The light bars give the statistical data for 1985 for the base case scenario. The dashed bars are the results of the energy conservation measures. The actions to reduce Btu's are reflected here; the amount of the dashed level above the 100 line is about half that of the light bar. The impact on the economy is minor if one accepts two percent differences in GNP and employment as minor. The impact on environmental emissions is more dramatic, at least for air emissions. For water effluents, little effect is noted, less than the economic change values.

SEAS USERS AND USES

Figure 15 presents a list of SEAS users and uses as well as available documentation of the Prototype system. A number of applications have been completed. Within EPA, support is underway to OPE, OSWMP and the Cost of Pollution Control Reports. Support of RTP and Water Programs with analyses is expected over the next few months. Outside EPA, major support and development collaboration are in progress with the groups listed in Figure 15.

The general SEAS documentation plan and products are presented in Figure 16. Each volume is complete by itself and fully documents SEAS information for a particular viewpoint.

SEAS IS A COLLECTION OF INDEPENDENTLY OPERATING MODELS AND ACCOUNTING PROGRAMS WHICH FORECAST THE FUTURE OF THE ECONOMY AND ITS ENVIRONMENTAL AND RESOURCE IMPACTS.

Figure 1
The Strategic Environmental Assessment System

PROVIDE A COMPREHENSIVE MEANS FOR EVALUATING THE LONG-RANGE IMPACT OF TRENDS, PROCESSES, ACTIVITIES, AND POLICIES ON THE ENVIRONMENT.

- NATIONAL/REGIONAL LEVEL
- MULTIMEDIA POLLUTION GENERATORS
- TEN- TO TWENTY-YEAR TIME HORIZON
- COMBINED MAN/MACHINE CAPABILITY
- ALTERNATIVE SCENARIOS
- STATE-OF-THE-ENVIRONMENT REPORT

Figure 2
SEAS Objectives

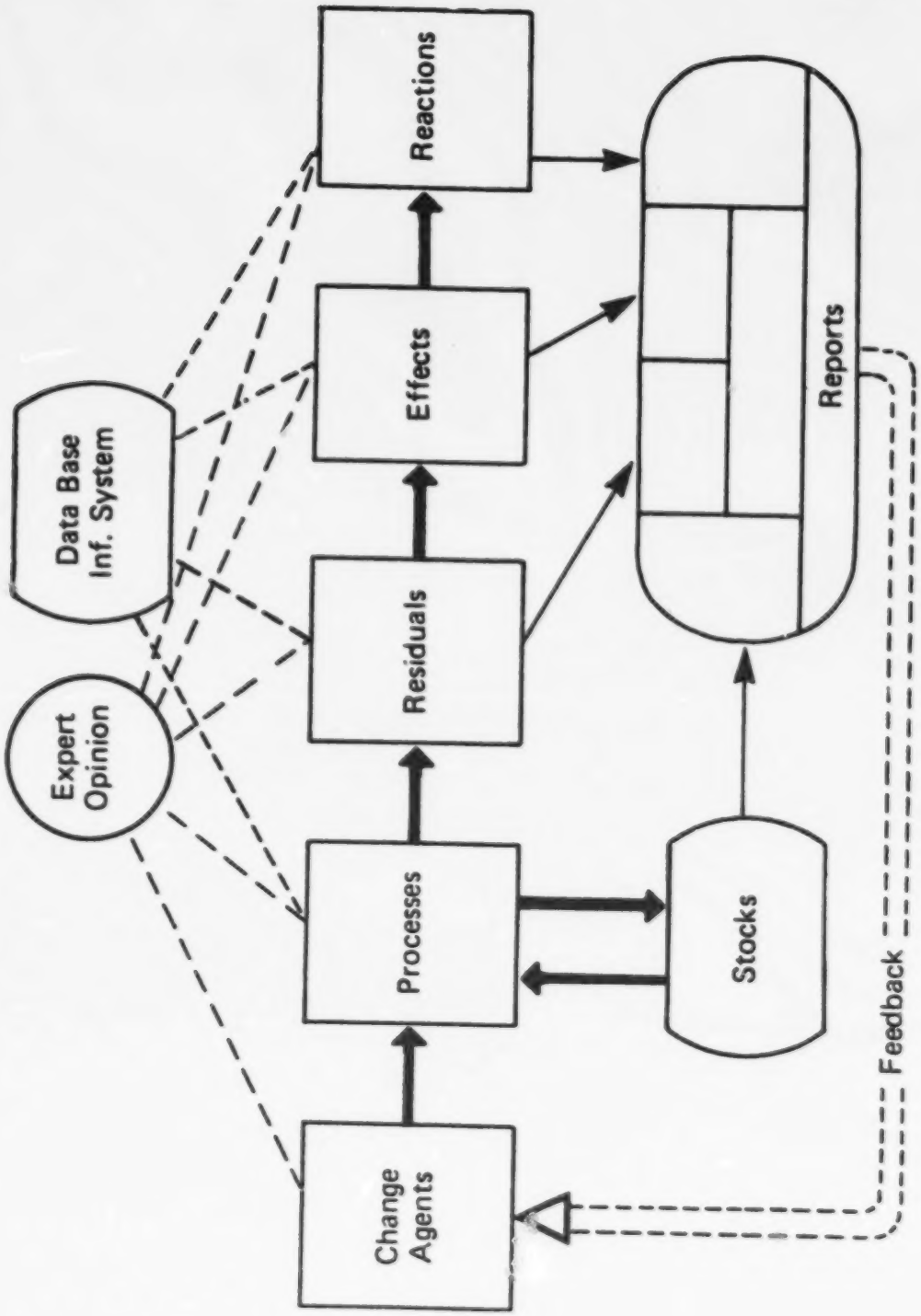


Figure 3
SEAS Concept

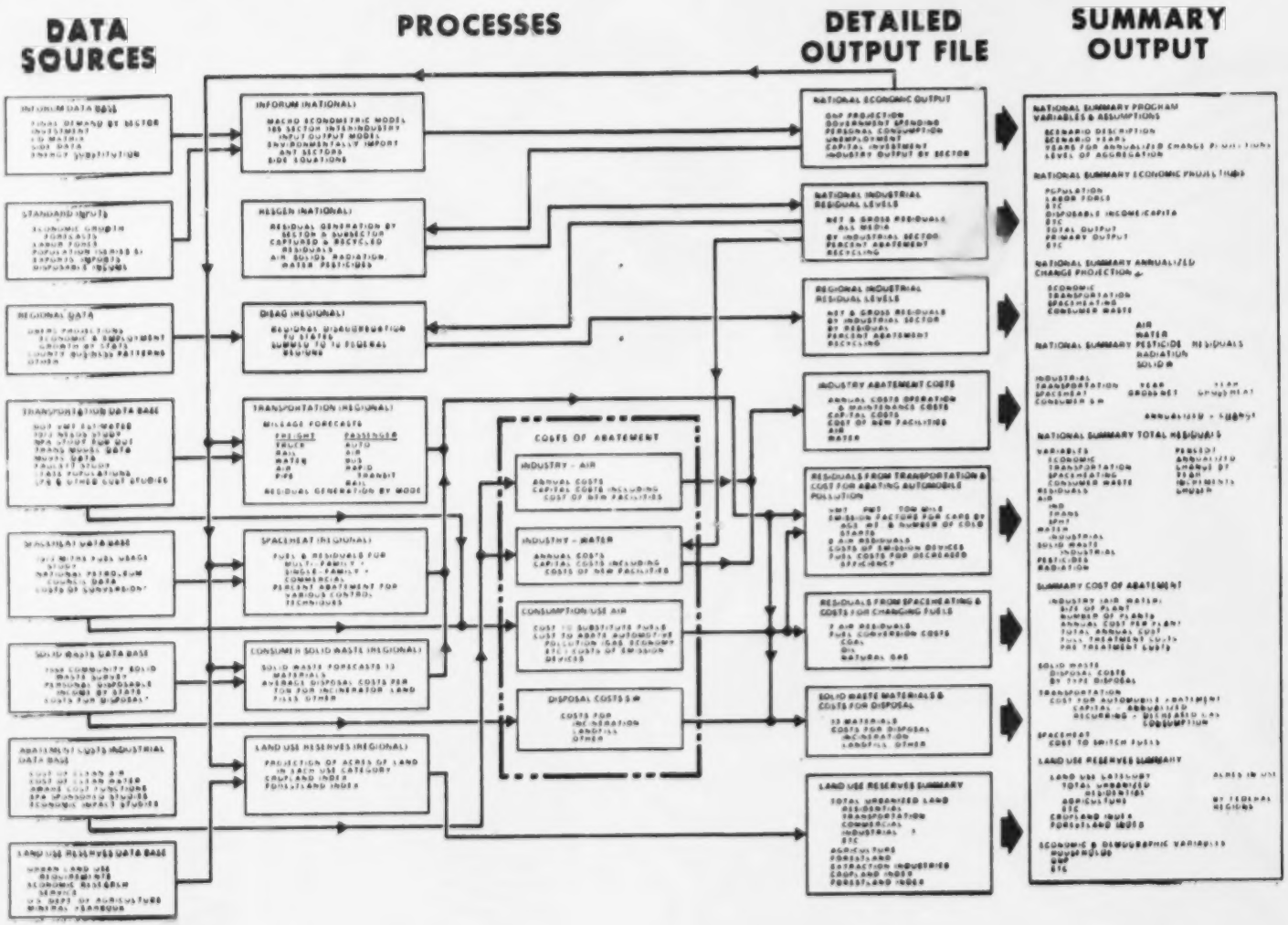
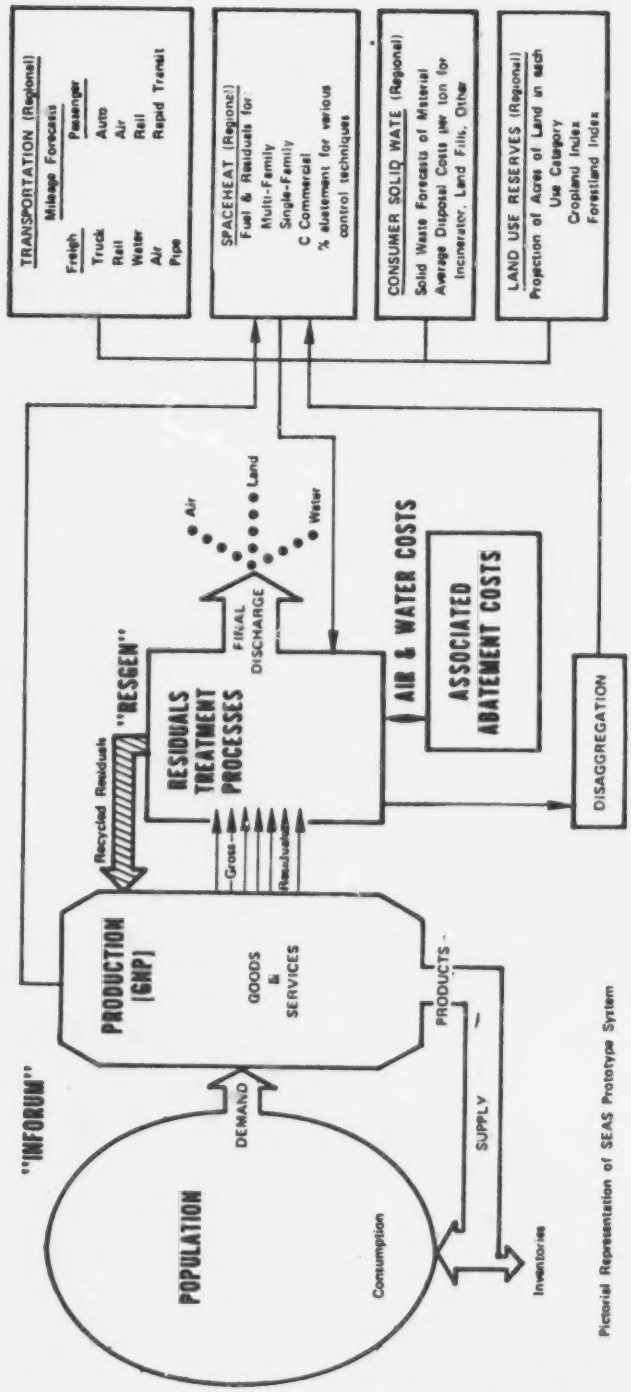


Figure 4
Strategic Environmental Assessment System

Figure 5
SEAS Prototype



Pictorial Representation of SEAS Prototype System

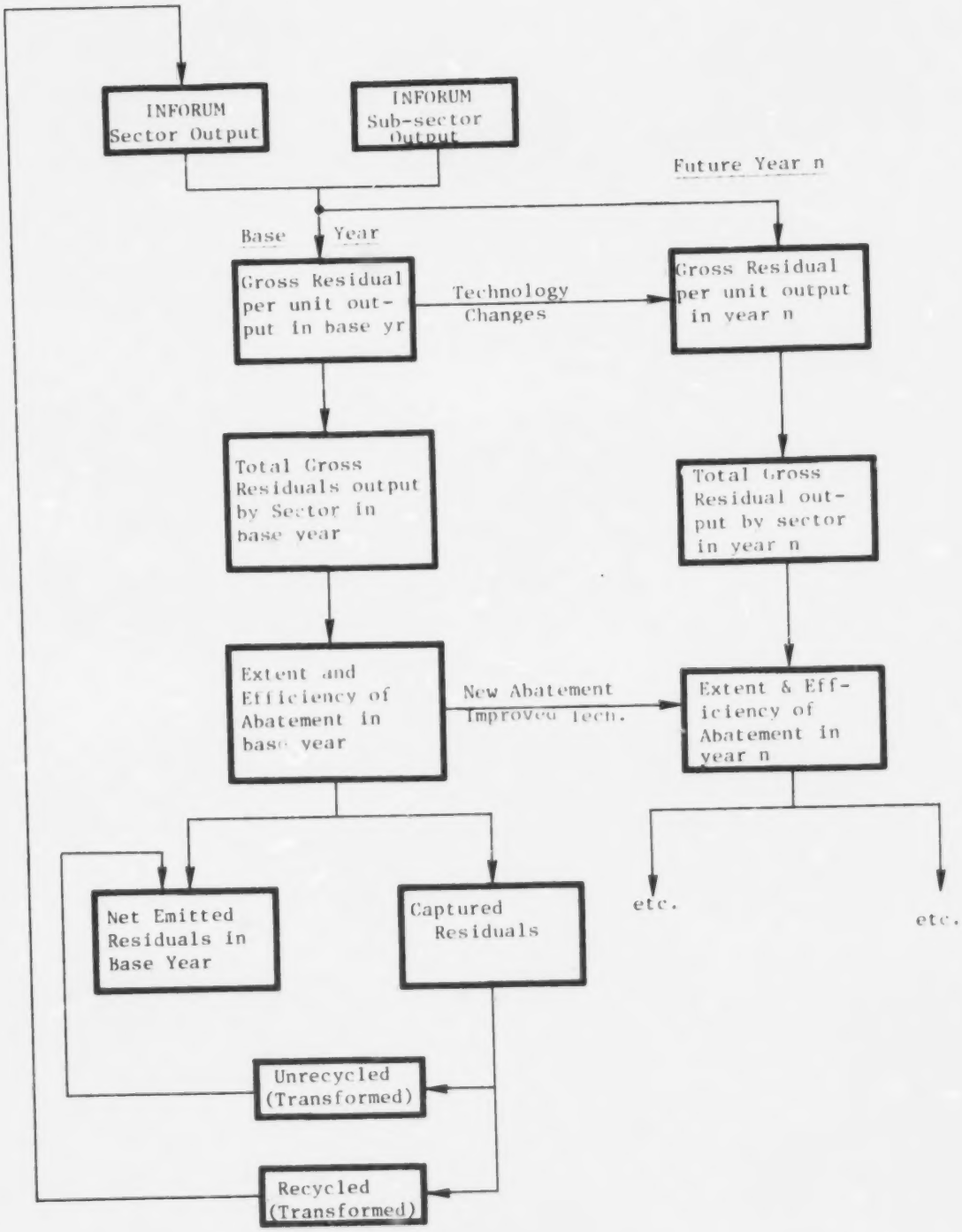


Figure 6
Residuals Framework

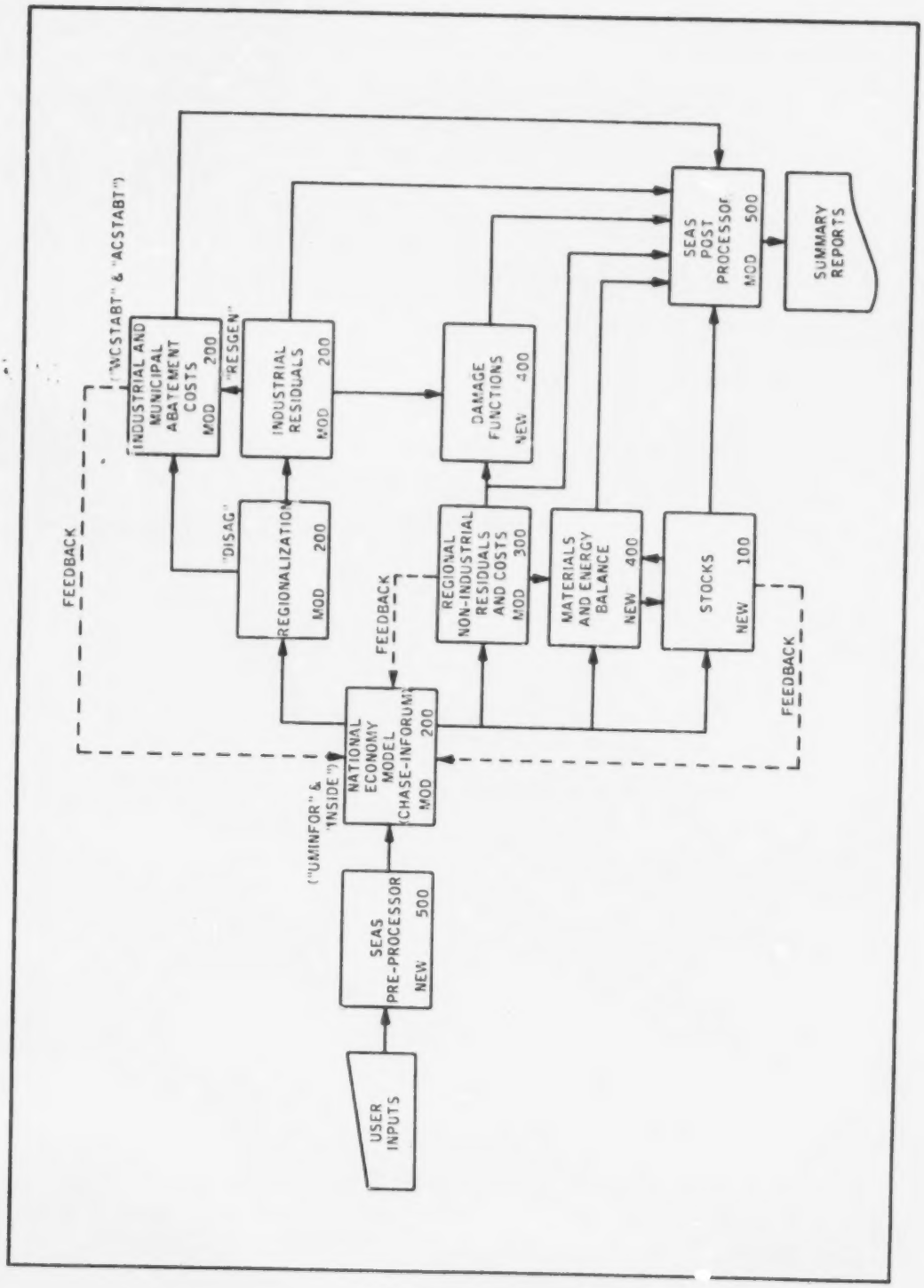
	Industrial/ Electrical	Transportation	Residential
AIR	TSP, SO_x, NO_x, CO, HC 10 particulates plus Mercury	TSP, SO_x, NO_x, CO, HC Lead	TSP, SO_x, NO_x, CO, HC
WATER	BOD, COD, SS, DS, two Nutrients, Phenols, 9 dissolved solids 6 others		
SOLID WASTE	6 types of combustible material 6 types of noncombustible material 2 types of mining waste		8 types of combustible 5 types of noncombustible material
PESTICIDES	Herbicides, Insecticides, Fungicides, Misc.		
RADIATION	4 Radionuclides to Air 8 Radionuclides to Water Radionuclides to Land		

Figure 7
Pollution by Major Sources

Program Name	Descriptive Title	Core Requirement (bytes)	Approximate CPU Time (MINS)
UMINFOR	Inter-Industry Forecasting Model (INFORUM, Modified)	315K	3.0
ENERGY	National Energy Forecasts	130K	0.5
INSIDE	Sector Disaggregation By Side Equations	130K	0.5
RESUPDT	RESGEN Coefficient Update	130K	0.3
RESCØPRT	RESGEN Coefficient Report	80K	.25
RESGEN	Generation of National Industrial Residuals	100K	1.5
RESPRT	RESGEN Report	105K	1.5
WCSTABT	Water Pollution Abatement Cost	100K	0.5
ACSTABT	Air Pollution Abatement Cost.	72K	0.5
DISAG	National Industrial Pollution Disaggregation	216K	13.0
DISAGPRT	National Industrial Pollution Disaggregation Report	100K	1.0
PTRANS	Passenger Transportation Residuals	126K	0.6
FTRANS	Freight Transportation Residuals	84K	0.4
SPACHEAT	Non-Industrial Space Heating Residuals	76K	0.3
CONRES	Consumer and Commercial Waste Residuals	75K	0.2
LANDUSE	Land-Use Reserves	100K	0.2
POSTCOMP	SEAS Post Processor	190K	2.5
SEAS			4

Figure 8
SEAS Program Characteristics

Figure 9
Phase III Overview of SEAS



ASSUMPTIONS:

ENVIRONMENTAL STANDARDS:

WATER PER CURRENT SCHEDULE

INDUSTRIAL AIR EMISSIONS FULLY ENFORCED IN 1977

AUTO EMISSIONS: INTERIM STANDARDS USED FOR 75-76 MODELS; FINAL STANDARDS USED FOR 77 MODELS

POPULATION:

- TOTAL, LABOR FORCE AND HOUSEHOLDS = SERIES E PROJECTION (DEPARTMENT OF COMMERCE)
- STATE SHARES SET BY OBERS

ECONOMIC:

- BASED ON PRE-ENERGY CRISIS DATA FOR: DISPOSABLE INCOME
  UNEMPLOYMENT LEVELS
- STATE SHARES SET BY OBERS

ENERGY USE/TECHNOLOGICAL CHANGE:

- PROJECTED FROM HISTORICAL TRENDS

Figure 10
The Basic Data Scenario

Parameter	Value 1971*	Growth Rate - Annual % 71-74	71-77	71-80
Air				
Particulates	20.50	-10.1%	-18.0%	-12.1%
SO_x	30.66	-1.3%	-7.2%	-4.5%
NO_x	19.91	+4.1%	+2.1%	+1.0%
HC	29.78	-3.2%	-4.8%	-4.7%
CO	125.70	-4.6%	-7.5%	-8.9%
Water				
BOD	8.13	-11.7%	-23.0%	-15.6%
COD	10.01	-11.4%	-21.7%	-14.7%
Suspended Solids	9.16	-11.9%	-22.8%	-15.5%
Dissolved Solids	25.30	-1.5%	-3.7%	- 2.2%
Nutrients	.06	-7.0%	-38.9%	-27.2%
Acids	.58	-11.9%	-15.5%	-10.3%
Bases	.11	-17.1%	-66.4%	-51.3%
Process Water	22.79	0.6%	-3.4%	-1.8%
Solid Waste				
Combustible	725.80	1.8%	2.2%	2.1%
Non-Combustible	63.03	4.1%	3.1%	2.6%
Mining	3316.19	3.4%	3.2%	3.0%
Pesticides	.136	1.8%	2.0%	2.0%
Radiation (Power Plants)				
Air	4578			
Water	7	41.8%	40.2%	34.2%
Land	1,150,000			

* Residuals in million tons except: Water in trillion gallons
Radiation in thousand curies

Figure 11
Summary: Pollution Data Output
(National - All Sources, Net)

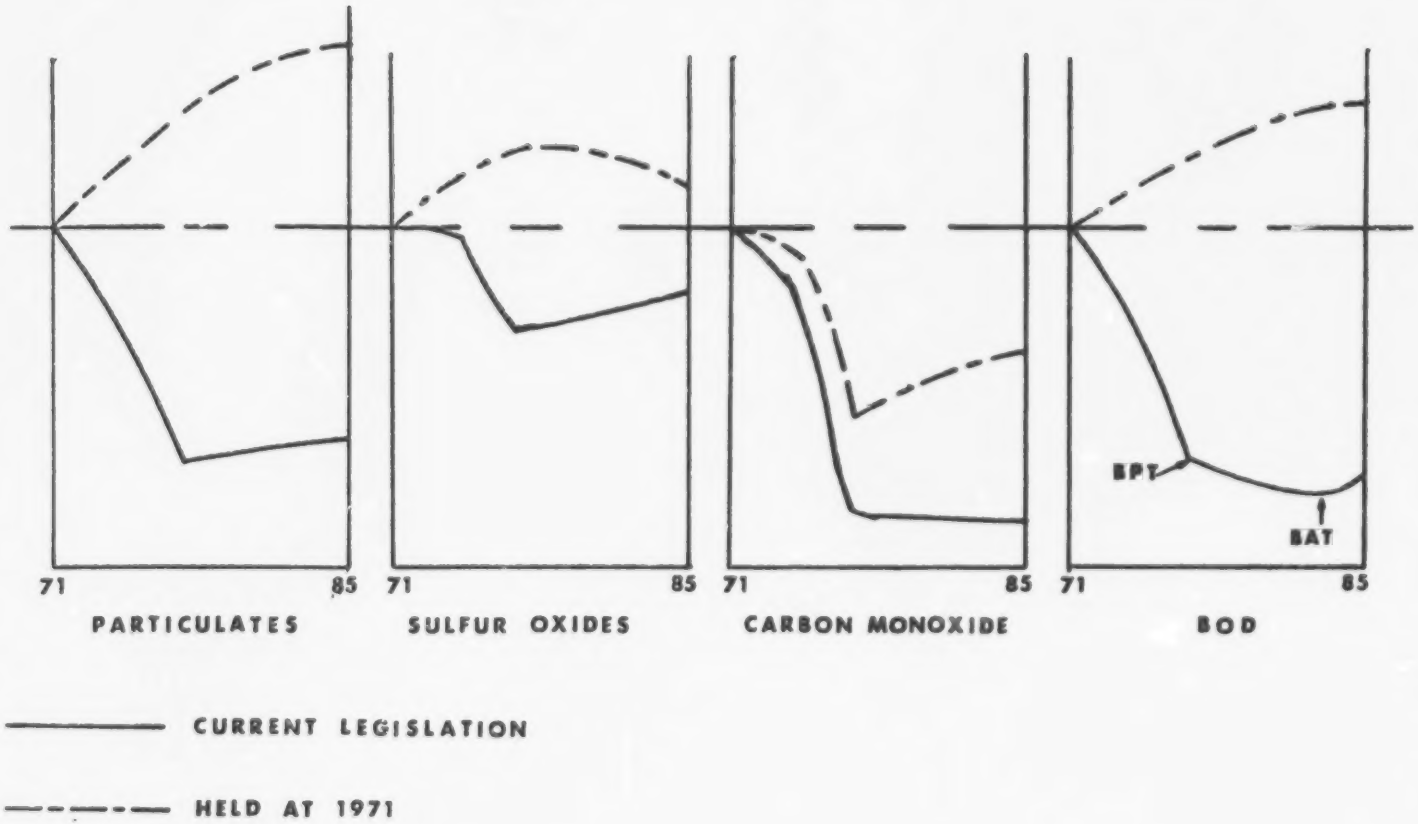
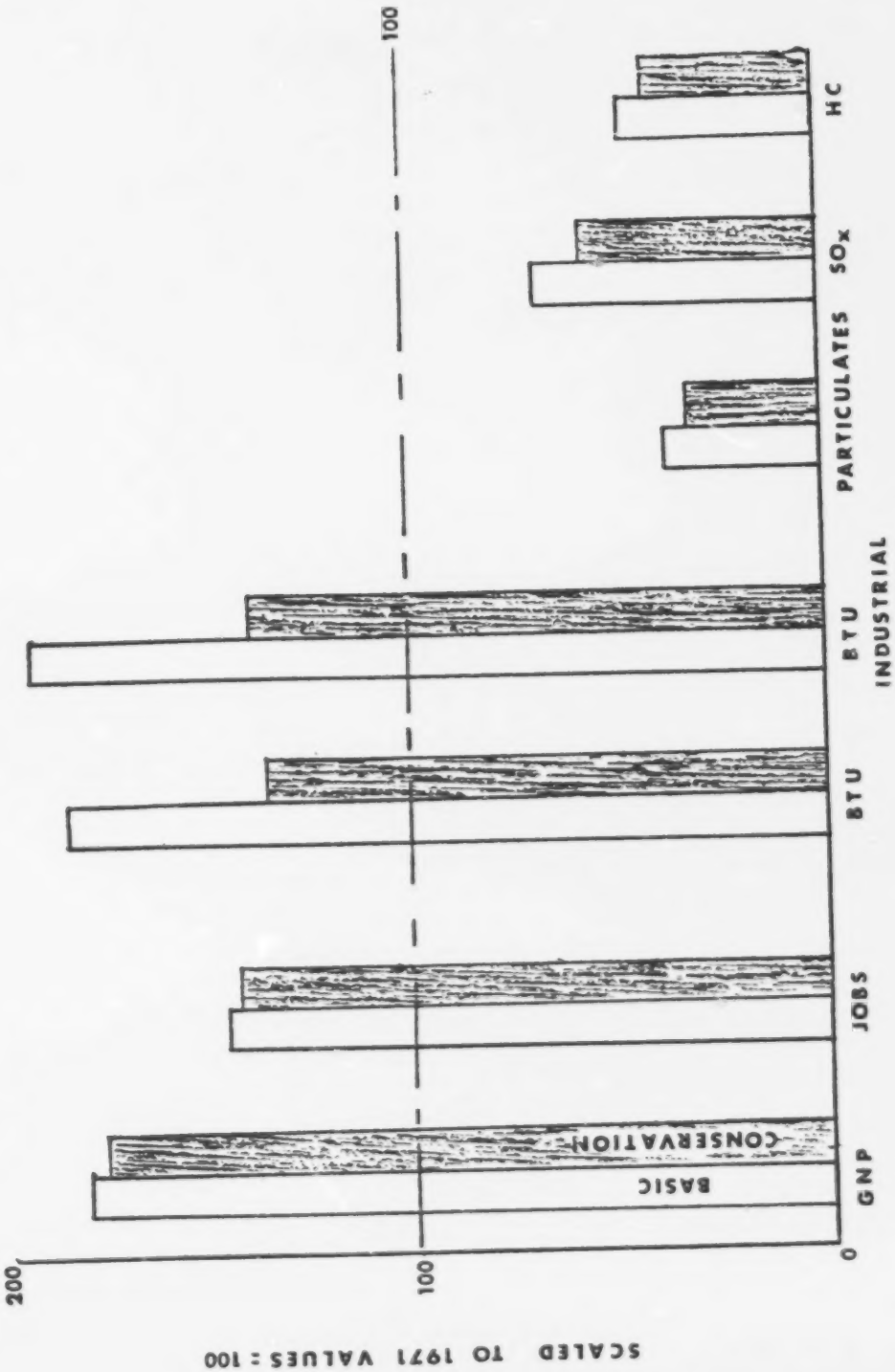


Figure 12
Pollution Control Laws

RESIDENTIAL	
Improved Insulation - savings of 25% in new housing (space conditioning & water heating - 30% of new housing stock)	1.23 quads
Old Stock Housing - 10% savings (comprise 70% of housing stock	1.15 quads
Substitution to Gas & Petroleum Products	2.54 quads
Reduction of non-all-electric housing stock by conversion to multi-family dwellings - 4.5% savings	.196 quads
COMMERCIAL	
Better Insulation - savings of 35%	1.76 quads
Retrofitting of existing buildings - savings 15%	1.76 quads
New Stock with total energy systems - savings of 30%	.50 quads
INDUSTRIAL	
Better Housekeeping - 25% savings	6.76 quads
Recycling & other Housekeeping measures	2.81 quads
TRANSPORTATION	
Lighter autos, less specific power, lower acceleration, improved driver procedures, reduction in non-essential driving caused by higher fuel costs - 35% savings	7.10 quads
Reduction of traffic by 1/2 from car-pooling	1.20 quads
Long-haul trucking replaced by train - 25% savings of fuel	.68 quads
Reduction of short-haul air travel by 40% by substitution of bus and rail - 35% savings	.90 quads

Figure 13
Conservation Measure

Figure 14
Energy Conservation Analysis, 1985



USERS AND USES

EPA

Cost of Clean Air, Water - 1975
AQMA Regional Definitions
Industrial Waste Stream Analysis
Energy Conservation Analysis

National Commission on Water Quality

Fall - 1974 Economic Analysis
Fall - 1975 Final Report

CEQ

Environmental Quality 1974 Report

Water Resources Council

Regional Water Effluent Forecast Data
Alternate Scenario Outcomes

NIH (RFF)

Long Range Economic/Environmental Projections
(Benefits Module)

Bureau of Labor Statistics

Environmental I/O Data
(Review of Productivity Data)

Figure 15
SEAS Users and Uses

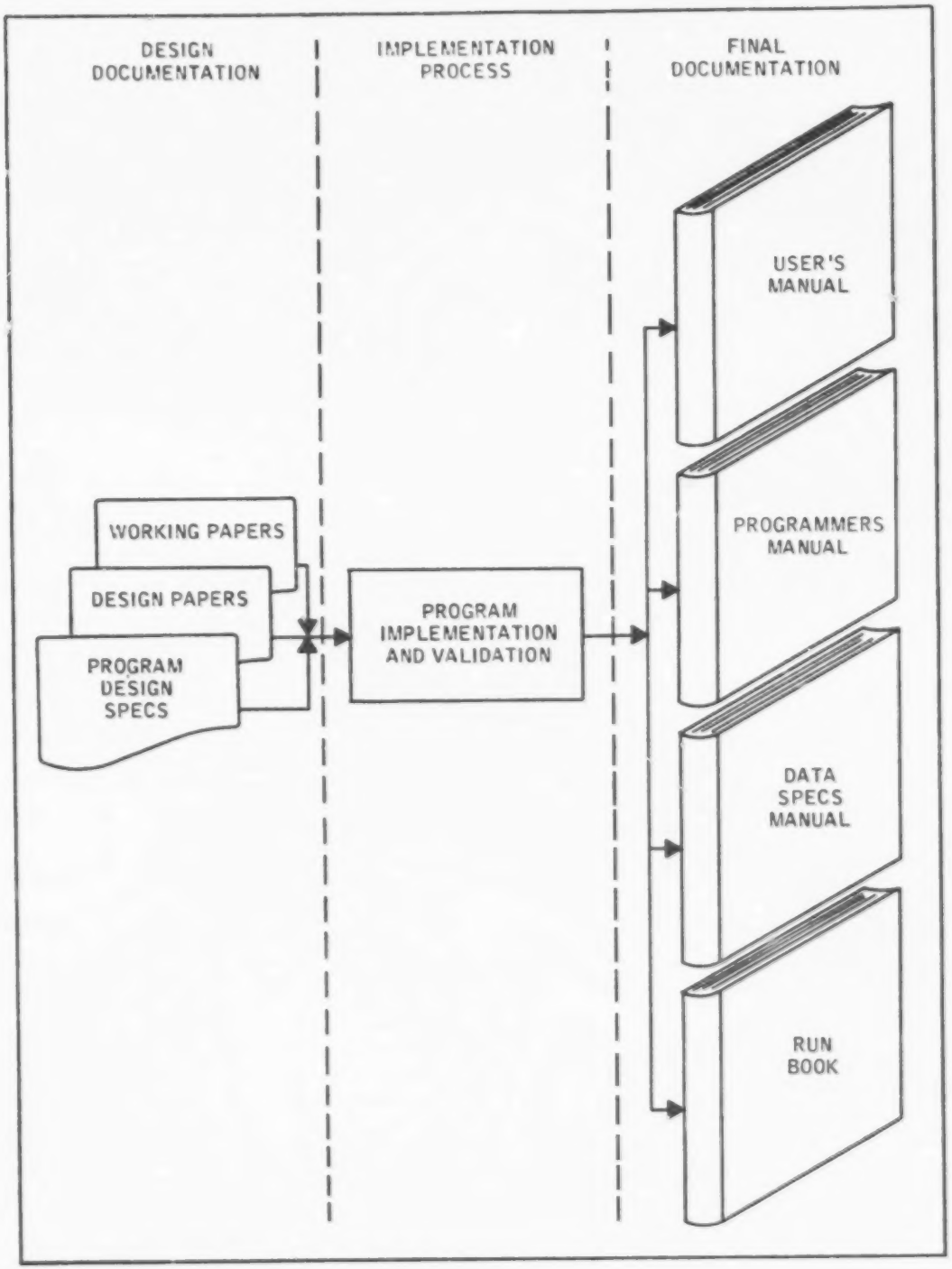


Figure 16
SEAS Documentation Plan

A COMPUTER CONTROLLED DATA ACQUISITION AND RETRIEVAL SYSTEM FOR USE IN AIR MONITORING PROGRAMS

By Marvin Hertz,* Kirby Kyle, Anne Duke, and George Ward

INTRODUCTION

Rockwell International Science Center and the Human Studies Laboratory of the Environmental Protection Agency (EPA), have designed and developed an automated data acquisition system for use in the Community Health Air Monitoring Program (CHAMP). The purpose of the CHAMP system is to acquire reliable air quality data for use in the Community Health Environmental Surveillance System (CHESS). The technical objective of the CHESS program is to correlate the quality of the air with the health of the population. An in-depth health effects study requires highly reliable data; the quantities of data required from each of the CHESS neighborhoods indicated need for automated acquisition, processing, and validation of the physical data.

To meet the program objectives, the CHAMP network will have 23 remote stations in six geographical areas around the country monitoring the air quality.

Changing program requirements and advances in aerometric instrumentation made flexibility in the acquisition system mandatory; capability for growth in the program suggested use of intelligent controllers at the remote stations. In order to avoid the loss of large quantities of data, rapid retrieval and diagnosis of data validity to alert remote station operators of failures or potential malfunctions in the instrumentation is required.

To satisfy various program objectives, a central computing facility at Research Triangle Park, North Carolina, with telephonic communication with each of the remote acquisitions systems and data retrieval capability was included in the program plan. A Digital Equipment Corporation (DEC) PDP-8 minicomputer was selected to control functions at the remote and the PDP-11/05 and PDP-11/40 computers operating in a dual processor mode were chosen for the central controller (Figure 1).

REMOTE

The PDP-8 minicomputer in the remote station serves as an interface between the analyzers and associated system, magnetic tape data storage, the remote field service operator, and the telecommunications network.

The PDP-8 data acquisition software package, DAX, handles all of the remote functions of acquisition and storage of data on magnetic tape, limited hardware control, interaction with the field operator and telecommunications. An innovation included in the software is the concept of an "instrument handler" to allow complete flexibility in controlling instrumentation with dissimilar operating characteristics. Each instrument handler is a sequence of code specifically pertaining to the functional peculiarities of the sensor; the handler's capabilities include actuation of appropriate valving, detection of conclusion of sensor cycle time, and observation of settling of a valve. The "instrument" is defined as the group of associated analog signals and status signals required for validation of the value detected from an aerometric sensor.

The analog signals are switched through a 48 channel multiplexor and digitized using a 41 microsec/word analog to digital converter. The analog signal from each channel is sampled once a second; values for each channel are averaged by the computer over a selectable period of time (nominally 5 minutes). In addition, 96 bits of digital information representing system status conditions are sampled. At the end of the averaging period, the detected voltage and status signals relevant to the validity of that datum are placed in a buffer in core. When the buffer fills, its contents are written to magnetic tape to await transmission on request from the central computer.

One of the several unique features of the CHAMP system is the extensive auxiliary information which is acquired and transmitted with the aerometric data. Of the 48 analog signals, 20 are presently allocated for direct ambient measurements (primary channels), and the remainder determine system operating conditions (secondary channels) (see Table I).

This information is transmitted to central for machine validation and is available to the field operator via local teletype to aid in daily maintenance of the

*Speaker

Table I
Analog Signals, Status Checks and Validity Criteria

Name	Instrument	Analog Slot No.	Validity Criteria
Wind speed	0	0	Limit check (0 volts would assume offset) 15 volt power supply
Wind direction	1	1	Limit check (0 volts ⇒ invalid) 15 volt power supply
Outside temperature	2	2	Limit check (0 volts ⇒ invalid) 15 volt power supply
Relative humidity	3	3	Limit check (0 volts ⇒ invalid) 15 volt power supply
Barometric pressure	4	4	Limit check (0 volts ⇒ invalid) 15 volt power supply
Inside temperature	5	14	Limit check (0 volts ⇒ invalid) 15 volt power supply
Particulate manifold impactor flow	6	16	Limit check (0 volts ⇒ invalid) Deviation from nominal
Particulate manifold filter flow	7	17	Limit check (0 volts ⇒ invalid)
Carbon monoxide	8	11	0 volts ⇒ invalid Range Calibrate Ambient = Power CO flow indicator Intake manifold Calibrate = Above + Instrument 13
Ozone	10	5	0 volts ⇒ invalid Range Calibrate Power Sample test point on Flow out of limits Sample manifold
Ozone flow		35	Sample flow in limits Mass flow test
Ethylene flow		33	Ethylene flow in limits Ethylene flow value Mass flow test Ozone air override Calibrate = Above + Instrument 13
PAN		18	PAN air flow PAN calibration value Sample manifold
Hydrogen flow		29	H_2 flow within limits PAN H_2 Mass flow test

Table I
Continued

Name	Instrument	Analog Slot No.	Validity Criteria
			0 volts ⇒ invalid
			Range
			Flame Out
Sulfur dioxide		9	Calibrate
			Power
			Sample flow test
			Sample manifold
			Flow within limits
SO_2 flow	11	39	Sample flow test
			Mass flow test
			Flow within limits
			H_2 flow off
H_2 flow		38	Mass flow test
			Flame override
			H_2 pressure out of limits
			Calibrate = Above + Instrument 12
NO		6	0 volts ⇒ invalid
			Range
NO_2		7	Range
NO_x		8	Range
		For all	Calibrate
			Power
			Sample test port - sample manifold
$NO\text{-}NO_x$ flow	12	36	Flow within limits
			Mass flow test
			Sample test port
			Flow within limits
Oxygen flow		37	Mass test value
			O_2 value
			O_2 pressure out of limits
Converter temperature		31	Within limits
Cell vacuum		45	Within limits
			$NO_x\text{-}NO/NO_2$ [illegible]
			Calibrate = Above + Instrument 13

Table I
Continued

Name	Instrument	Analog Slot No.	Validity Criteria
Dilution air - Range 1	13	41	0 volts ⇒ invalid
Dilution air - Range 2		42	All Word 7
Dilution air - Range 3		43	Heatless air dryer
			Air pressure out of limits
			Humidifier
Calibration flow		34	NO on
			H_2S on
			CO on
			Mass flow test
			Gas to manifold on
Oxygen flow		40	O_3 flow measure
			NO_2 flow measure
			SO_2 flow measure
			Mass flow test
			SO_2 span gas
			NO_2 span gas
			O_3 span gas
Ozone generator		28	O_3 generator on
SO_2 permeation - tube temperature		44	Within limits
Combustion alarm	14	46	Within limits
			LEL
			Exhaust manifold
			24 volt power supply
			Illegal entry
			H_2 ext
Reference voltage		32	Within limits

instrumentation. The local operator may include supplementary information as "manual entry data" or "journal entries"; this information is recorded on magnetic tape and transmitted to central as separate data for use in the quality assurance reports and area engineer data validation.

Since each instrument may have a different averaging or cycle time, the capability for asynchronous handling had to be implemented. An interrupt is generated by the system clock at a fixed rate and an instantaneous voltage is acquired from each sensor; the handler for each on-line instrument is called in the acquisition mode and all necessary operations of control, acquisition, and storage are performed. The instrument handler may also be called in three other modes: 1) the initialize mode which allows the field operator to specify operating parameters; 2) the reset mode which resets cycles and initializes the time sequencing; and 3) the calibrate mode which allows the field operator to perform the calibration procedures interactively through the console teletype.

The DAX program is organized around a priority interrupt scheme coupled to foreground/background mode of operation; data acquisition occurs at the highest priority level to guarantee no loss of data. Background functions are handled as processor time is available.

DAX handles the data transmission task in foreground mode. When polled from central, the magnetic tape is backed up to the position on the tape where the last transmission terminated. Data is read into core and transmitted record by record with an acknowledgement protocol at the conclusion of the transmission of each record.

TELECOMMUNICATIONS

Data is retrieved on the request of the central computer system from each of the remote stations via a dial up phone line at two-hour intervals. The central and remote computers converse via a telecommunications system consisting of modems operating in full duplex mode at the rate of 1200 baud from remote to central and 150 baud in the reverse direction. Polling is under complete control of the central controller and is the primary function of one of the two central processors. A file on disk contains the phone number of each station in the format required by the Automatic Calling Unit (ACU) interfaced to the PDP-11 system. An alterable polling sequence queue is also disk resident. A rigid protocol has been established to guarantee accurate transmission and retrieval of data. Central makes several tries to establish contact with a remote station before abandoning the attempt and placing the station at the bottom of the polling queue. A hardware carrier detect protocol establishes the communications link. Each frame is checked for parity and framing errors by the modem controller. Checksums are computed for each 512 frame record and compared by the computer. An acknowledge character is exchanged indicating correct receipt of the record. Should any of the tests fail, several transmission retries are made. Communications are terminated by receipt of a character from the remote indicating the end of data or by failure of the remote to transmit in the requisite period of time.

CENTRAL CONTROLLER

The central controller for the CHAMP network is a dual processor system with a full complement of input, storage, and display peripherals. The heavy burden on processor time placed by the telecommunications and real-time processing of the large quantities of data anticipated justified the choice of the dual processor system. A PDP-11/40 was selected to perform the tasks associated with management of the large data base to be generated by the network. The telecommunications and real-time processing tasks are handled by a PDP-11/05 computer. The two processors are interconnected by a device called the UNIBUS window which takes advantage of the unified asynchronous data path architecture of the 11 system. The window allows each processor to address core and peripherals on the other processor as if it were its own. In addition, the DEC Memory Management hardware option was added to the PDP-11/40 to handle addressing above 32K in the 16 bit system. An extensive complement of peripherals including two cartridge type disks, three tape drives, an electrostatic printer-plotter, line printer, and CRT display were initially selected; the rapid retrieval requirements for large quantities of historical data suggested addition of a 2316 type disk file and implementation of an extensive file management system after the start of the program.

One of the program requirements is timely diagnosis of system faults at central; the need to perform a cursory check on all data as received was apparent. Hence, the PDP-11/05 processor was dedicated to the telecommunications and real-time checking of values against preset emergency and instrument malfunction limits, and storing of data on a magnetic tape as received. The Level 1 tapes so created are an image of the tapes recorded at the remote station, but with the data interleaved by station in the polling order. The 11/40

processor is devoted primarily to data validation and quality assurance tasks. Although the DEC supplied Disk Operating System (DOS), is suited to data reduction, analysis and display, the real-time applications and problem of shared peripherals required additional supervisory software. A Real-Time Monitoring System (RTMS), developed as a Rockwell proprietary package, is used in the CHAMP network to allow the two processors to act in a single foreground/background mode. RTMS controls peripherals access, implements memory management using the core in the 11/40 not accessible by DOS, and protects the core memory allotted to each task from destruction by other tasks. RTMS supports the dual modes of operation of the 11 system, kernel and user; RTMS in the 11/40 runs the kernel mode for system protection with the nonreal-time work in the background under DOS running in user mode.

A fundamental system objective is to provide machine validation of data to ensure reliability (Figure 2). Each primary channel datum is checked for nonzero voltage, normal setting of all valves, power to the instrument and digitally measured tolerances for proper ambient sampling. Related secondary channel analog signals are tested to verify that values are within tolerances and status conditions necessary for correct sampling are checked. Information referring to the general system operation is checked on a five-minute basis.

Once a datum is determined by the machine to be valid, the engineering unit value is computed by applying the calibration constants to the voltage; this value is stored on tape and disk. This Level 2 (validated) data is used in further processing and, when accepted by the area engineering, is the end product of the CHAMP system.

Since a crucial factor in obtaining accurate data is the calibration of the instrumentation, the same in-depth validation procedure is applied to the calibration operation to verify proper flows from the calibration gases and correct valve setting on the flow system. If an improper calibration is detected at central, the field operator is contacted and advised to repeat the operation. The calibration constants are computed from the known concentrations of the source gases; the constants are checked for goodness of fit to the measured points and for linearity where applicable.

Initially, the validation procedure was envisioned as a daily procedure involving processing of the Level 1 tapes from the previous day to create a Level 2 tape containing only machine validated data. Addition of the large random access disk impacted the procedure for validation; data will be recorded on the disk as polled. It became apparent that validation could be done as a batch operation at regular intervals or on demand. Hence, a user could have access to validated data at any time; tape need be used only as a secondary storage medium for archival or temporary storage purposes. The validation procedure will reveal system problems and alert cognizant personnel at central; hence, more frequent validation will contribute to timely anticipation and prevention or correction of system difficulties.

Several additional features have been included in the background software to assure prompt diagnosis of system problems at central. The system status condition and analog values may be displayed on the CRT as each station is polled. Secondly, a user may "demand" poll or alter the polling sequence to retrieve all data from a particular station since the last polling. Thirdly, a user at the central facility may "view" the real-time operation of a remote station. The one-second values, as sampled at a remote station, may be transmitted to the central facility and displayed on the CRT within the limitations of the transmission rate.

A quality assurance file is maintained on a separate disk unit. The maintenance operations performed at each remote site are encoded and entered as "manual entry" data, transmitted to central and recorded in this file during validation procedure. Data acquired during the calibration procedure is saved. Plots of the calibration values over an extended period of time may indicate a steady degradation in a sensor which would not be detected during the normal validation procedure. Correlations between parameters or stations may indicate a problem which could escape the validation tests. Tabulations and displays of the above, in addition to reports of failures and other anomalies, are printed on a regular basis and are available to the user on demand.

A statistical package which fully utilizes the peripheral equipment will be available to users. The program will collect the requested parameters from the data base and perform the statistical summary desired. Included in the repertoire will be calculation of frequency distributions, correlations, threshold exceedances, and minima, maxima and mean values. Reports may be tabulated or plotted; a software goal is to make all summaries device independent, allowing the user to select the output media at run time. The statistical subroutines will be of particular value to the field manager, quality assurance personnel, and the area engineers at EPA.

Final acceptance of validated data is the responsibility of the area engineer. Data editing capability is provided prior to the production of the final data base tape for the health effects study. It is anticipated that his subjective evaluation and verification of data quality will entail heavy usage of the statistical and display tools available to him.

The features of the automated CHAMP system represent a significant advance over previous systems, both in terms of quantity and reliability of data.

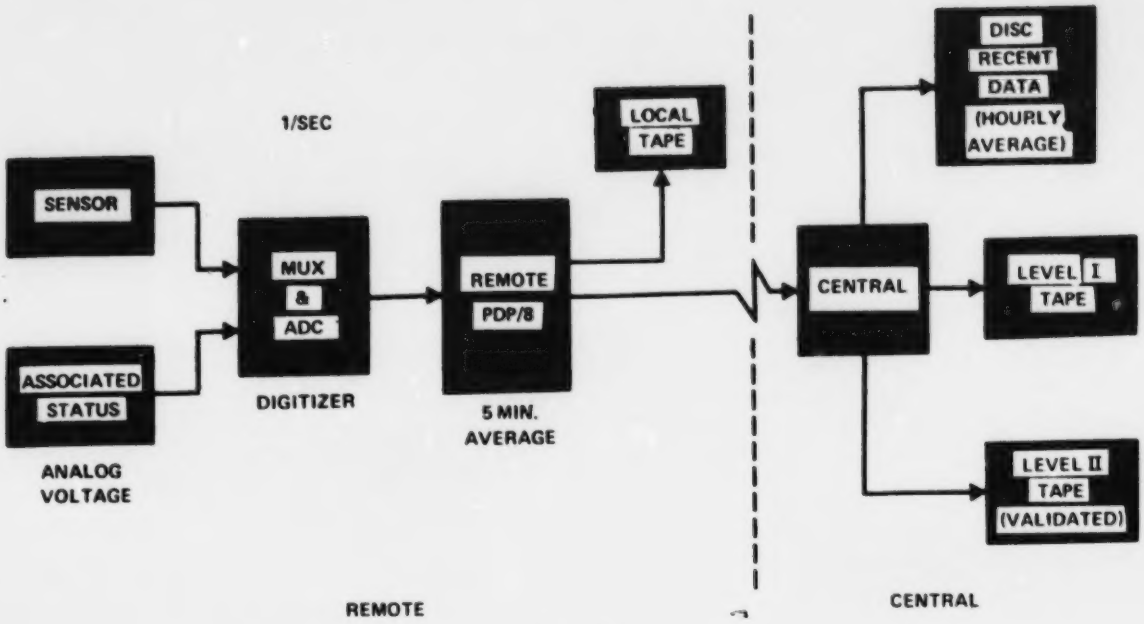


Figure 1
CHAMP Data Flow

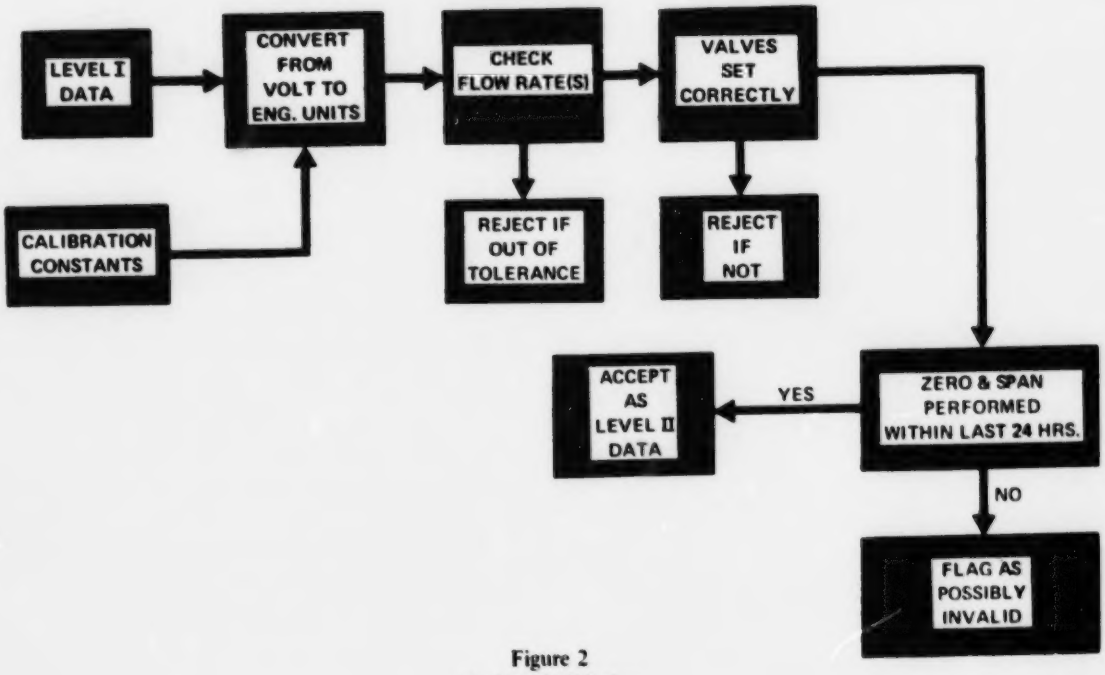


Figure 2
Machine Validation

AN AQUATIC ECOSYSTEM SIMULATOR AND VS/8

By David M. Cline

INTRODUCTION

The Aquatic Ecosystem Simulator (AEcoS) is a large-scale simulation chamber in which a variety of environmental parameters may be controlled and monitored over a wide dynamic range. AEcoS will be utilized to verify aquatic ecosystem mathematical models with biological experiments conducted under controlled conditions. This approach will provide insight into aquatic ecosystems that cannot be found in bench experiments or in the real world.

A minicomputer system was chosen to control and monitor the AEcoS, as well as to process the acquired information. The system has since evolved from a minimum configuration, which was specified to perform calibration of the many transducers, to a fairly sophisticated computer system that allows full control of the environmental parameters, data acquisition of transducers as well as analytical instruments, and background processing.

Considerable attention was given to the area of background processing and, as a result, a real-time software operating system was developed to allow multiple independent background tasks to execute in a virtual memory mode. Since the software was not available from the manufacturer, a decision was made to implement the real-time virtual memory operating system in-house.

AQUATIC ECOSYSTEM SIMULATOR

AEcoS is the implementation of an idea spawned by Dr. Walter M. Sanders, III, Chief, Freshwater Ecosystem Branch, Southeast Environmental Research Laboratory, National Environmental Research Center, Corvallis, Oregon. The concept, initially conceived in the early 1960's has been transformed into a completed facility that was formally dedicated in March 1973. Briefly, AEcoS is a chamber 72 feet long, 12 feet wide, and 9 feet high and houses an experimental stream 64 feet long, 1.5 feet wide, and 2 feet deep. The artificial stream consists of eight 8-foot Teflon lined Plexiglass sections; thus, any of the 8-foot sections may be altered without altering the other sections.

Air and influent water temperature can be controlled with an accuracy of one-half of 1°C over the full dynamic range of 0° to 40°C. Independent of temperature, relative humidity can be controlled to within 2 percent over the dynamic range of 20 to 95 percent and can be changed by a maximum of 60 percent in one hour.

The influent water supply can provide up to a maximum of 2000 gallons per day through the use of four 30-gallon per hour stainless steel deionizer-distillation units, each of which has a storage capacity of 500 gallons. Dissolved gases, nutrients, and pollutants may be introduced into the water for preparation of a tailored experimental design.

Radiant energy simulation within the chamber is accomplished by combining lamps of different colors and intensities. The radiant energy system consists of a light feedback control system, 833 florescent lamps, and 100 infrared lamps, all of which consume a maximum of 200 kilowatts.

Instrumentation for AEcoS includes 48 temperature sensors at various depths along the water channel. Sensors are also installed at each eight-foot section to measure pH. Nine chemical analyses are performed at each of nine locations by hydraulically multiplexing fluid samples with a series of solenoid operated valves and routing them into only nine Technicon autoanalyzers.

COMPUTER HARDWARE

A Digital Equipment Corporation PDP-8/E minicomputer was specified initially in lieu of hard-wired logic to act as a calibration system for the AEcoS instrumentation system. The decision was based on cost effectiveness and flexibility of the minicomputer versus hard-wired logic. Subsequently, a management level decision was made to upgrade the minicomputer in order to provide control of the environmental parameters (air temperature, water temperature, and relative humidity), the water flow, and the radiant energy system. In addition, the capability to perform preliminary data analysis and the capability of exporting the acquired data in a form compatible with EPA's computer facilities were required.

To this end, the following hardware was added to upgrade the minicomputer:

- Hardware multiply/divide unit
- 12K of core memory
- Power fail/auto restart
- 128 analog inputs
- 4 digital-to-analog converters
- 5 digital input/output units
- 9-track industry compatible magnetic tape drive
- 1.6 million word disk cartridge unit.

COMPUTER SOFTWARE

In conjunction with the decision to upgrade the minicomputer hardware, a decision was also made to upgrade the software for the minicomputer system.

The initial real-time software executive used with the AEcoS computer system was obtained from the Oak Ridge National Laboratory (ORNL), which had developed the software for use by many of its minicomputer systems. The software was core resident and supported a single teletypewriter, a real-time clock, and a maximum of 4K of core memory. The executive supported the scheduling and execution of seven periodic tasks in a strict priority manner.

Since the executive was coded for a predecessor of the PDP-8/E, several modifications were necessary to take advantage of the available hardware. Using the initial scheduling algorithm as a base, algorithms were modified or added to provide support for a real-time programmable clock, a digital I/O unit, an extended arithmetic element, an industry compatible magnetic tape unit, and an analog multiplexer. The resulting real-time executive has performed satisfactorily and is currently being used on another SERL computer system.

Also, a disk subsystem and additional memory were obtained for the storage of data and programs, and the autoanalyzers and the gas chromatographs were interfaced to the AEcoS computer system. Computer terminal support was also desired in the AEcoS control room and in the AEcoS chemical support laboratory to permit small computer programs to be executed on demand. A conceptual design of a new real-time virtual memory operating system was therefore defined.

The resulting specifications required that the ability to perform program preparation, assembling or compiling, program loading (absolute or relocatable), program execution, device independent I/O, memory management of 32K core memory, and file management be provided while real-time tasks were being scheduled and executed. It was also specified that these requirements be fulfilled at each of the seven teletypewriters supported by the system, which would demand a PDP-8/E with a total of 224K of memory exclusive of the monitor and real-time task memory requirements. However, the PDP-8/E can only be expanded to 32K of memory. An investigation of the hardware revealed that the minicomputer could be made to interrupt the software executive system whenever a hardware request was made to change from one memory field to another. Use of this fact made the implementation of the virtual memory system (VS/8) possible while maintaining software compatibility with existing PDP-8 programs.

Since the implementation of the entire VS/8 system would be a time-consuming effort, a working system, without the device independence and file management attribute, was developed to provide the capability of using a higher level language for background processing. The hardware manufacturer provided an interpretative language, with its own program editor, which was modified to make use of the VS/8 supervisory calls. The resulting software was utilized immediately to provide interaction with the chemists in the chemical analysis laboratory in conjunction with the automation of the analytical instruments, and to provide graphical representation of the AEcoS system parameters in real time. This version of VS/8 is release 3.0 and will be used and supported until release 4.0 is completed.

The capabilities of VS/8, release 4.0 are as follows:

1. The system permits 8 levels of priority scheduling. Within each level n tasks (where n is dynamically alterable) are executed in a round robin fashion. All tasks at any given priority will execute before any tasks of lower priority.

2. A virtual task has 8 core fields (32K) unique to that task, which correspond logically to the physical fields 0-7 of a hypothetical 32K machine. Greater region sizes would be possible by sacrificing direct compatibility with a nonvirtual PDP 8/E.

3. Any given task may have a maximum of 4 files open simultaneously within a system-wide maximum of 32 opened files. Files shared by several tasks count as one open file for the system limitation. Files are identified by a device name, a file number or file name, and account name dependent on the type of device on which they reside. Direct access files may specify optional protection features such as write lock, and may select the degree of access permitted to other users.

4. Supported devices are subdivided into classes and are too numerous to list but include essentially any I/O device that may be interfaced to the hardware.

The system has the following requirements:

1. Minimum configuration for a solely real-time system includes 12K of memory, any real-time clock, and appropriate I/O and control devices.

2. Minimum configuration for a solely virtual system includes 16K of memory, any real-time clock, at least 256K disk memory, and one to seven terminal devices.

3. Minimum combined system includes the required devices for the stand alone systems with 20K of memory.

CONCLUSION

The viable simulation system presented includes the simulator and a support computer system. Without the computer system, the AEcoS would be difficult, if not impossible, to utilize because of the problems inherent in manual acquisition of data, control of the chamber, and analysis of collected data. With in-house manpower, a hardware/software computer system has been developed which not only supports AEcoS, but also provides a small in-house computer capability.

AN ON-LINE PDP-12 SPIROMETRY PACKAGE THAT SUPPORTS BOTH ONSITE AND REMOTE STATIONS

By Sam D. Bryan

PROJECT BACKGROUND

The project described below is an ongoing project of the Clinical Studies Branch in the Human Studies Laboratory at the Research Triangle Park, North Carolina NERC. The mission of the Clinical Studies Branch is to study the effects of pollution, primarily air pollution, on human health in a clinical laboratory setting where quantitative measurements are made.

The lung is one of the main organs of interest. One of the lung projects began about two years ago when the branch set out to determine what effect air pollution might have on the growth of lung function in children. An answer to this implied that the pattern of lung growth in a clean air environment first be determined, a requirement that led to the computer system described here. A suitable population of preteenage children was located at a school in Chapel Hill, North Carolin.

Although many other lung function tests exist, the investigators chose to concentrate on spirometry tests. Spirometry tests measure how effective the lung is simply as a machine that moves air back and forth. The measures relate largely to how much air the lung can store and how fast it can move air during different stages of the respiratory cycle. The subjects are exhorted by the attending technician to perform at their best effort; sometimes many trials are required. For a detailed description of spirometry, see Comroe.*

The next question was how to carry out the measurements. The project team had access to a PDP-12 computer, which they installed at the school, and located a PDP-12 spirometry package developed at the University of Chicago Biomedical Computation Facility by Drs. Earle and Domizi.† Digital Equipment Corporation (DEC) is marketing a PDP-8 system based on the University of Chicago package, which they call the Pulmonary Testing System.†† The initial task was to modify the package for the project's configuration and to redesign it for use within our experimental setting, since its original use was within a patient care setting. Several measurements that were of particular interest were also added.

The team rejected the approach of performing the measurements in the old manual way since there was an automated system available to use more or less off-the-shelf. The automated approach is more accurate and many times faster than the manual way. Speed was especially important since the children involved in the study would often have to repeat the breathing trials many times in order to produce usable results. The on-line approach made it possible to know within seconds if further testing was required.

COMPUTER CONFIGURATION

In relation to other PDP-12's, the project team has a rather large configuration. It has 16K of core, floating-point processor, 1.2-million word cartridge disk, 9-track industry compatible magnetic tape, 300 line-per-minute line printer, flatbed plotter, as well as the usual LINC-tapes, CRT scope, A/D converter and relays.

The configuration required for the spirometry package is considerably smaller and is shown in Figure 1. The programs run in 8K and use the LINC tapes for program and data storage. Team members can copy the LINC tape data to 9-track tape for processing on a large-scale computer such as the Univac 1110.

MEDICAL INSTRUMENTATION

The transducer used to convert the rate of flow of the subject's breath into an electrical signal is the Medscience 465 Hi-Fi spirometer. The subject breathes into a tube which is connected to a bellows whose back end moves back and forth on a track as the subject breathes out and in. A potentiometer is moved in the process and,

*J.H. Comroe *et al.*, *The Lung: Clinical Physiology and Pulmonary Function Tests*, 2nd ed. (Chicago: Year Book Medical Publishers, Inc., 1963)

†D.B. Domizi and R.H. Earle, "On-Line Pulmonary Function Analysis: Program Design," *DECUS Proceedings* (Fall 1970).

††Digital Equipment Corporation, *Pulmonary Testing Systems: Spirometry* (Maynard, Massachesetts, Digital Equipment Corporation, 1973).

in turn, serves as part of a resistor network in an amplifier that produces an output voltage proportional to flow. This voltage is sampled by the computer's A/D converter 80 times per second under control of a programmable clock. The analog sample in the range ±1 volt is converted into a digit in the range ±511.

EXPERIMENTAL PROTOCOL AND PROGRAM LOGIC

Before subject testing is begun, the system must be calibrated so that the computer correctly interprets the digital values it receives from the A/D converter in terms of the corresponding flow rate of air in the spirometer. By using the teletype, the technician informs the computer of the occurrence of a zero flow (baseline) and of a 10-liter/second flow calibration signal provided by a special amplifier circuit. The two digital values observed can then be used by the program in a simple interpolation formula to relate any observed digital value to the corresponding flow rate. For example, assume that the baseline signal yields a digitized value of 100 and that the 10-liter/second flow calibration signal yields a digitized value of 500. Then, if a value of 300 is observed, we would know that it was caused by a 5-liter/second positive flow in the spirometer.

After calibration the system is ready for subject testing. The subject is "logged in" by the technician, who uses the teletype to provide certain identifying information that will subsequently be recorded on LINC tape along with the digitized flow data. The identifying information includes the date, the subject's ID code, his height, weight and a symptom code if he has a cold, cough, etc.

The subject is then instructed to place his mouth over the mouthpiece leading to the spirometer, to take a deep breath, and then to breathe out as quickly and completely as possible. The flow signal is monitored on the CRT screen as a left-to-right oscilloscope type sweep. If the subject appears not to be performing the breathing maneuver properly, he is encouraged to do better. Once a successful maneuver has been completed, as observed on the CRT, the technician strikes a key on the teletype and thereby causes the computer to analyze the last 10 seconds of flow data. The first part of the analysis entails the detection of several events of the respiratory cycle, including beginning and end of inspiration and beginning and end of expiration. These events (points in time) are then indicated on the scope as hash marks superimposed on the analyzed flow signal. If the technician sees that the computer has made the wrong determination, the analysis is rejected and the subject is instructed to begin a new trial.

Once the pattern recognition part is performed successfully, the second stage of analysis, the measurement stage, is begun. One measurement, for example, is the maximum expiratory flow rate. Since the last 10 seconds of flow data are stored in core and since the computer knows the beginning and end expiration points, a simple search of the corresponding part of the data storage buffer yields the maximum expired flow value and the time it occurred. To compute the capacity (technically speaking the "forced vital capacity") measure of the lung, the flow values of the expiration part of the data buffer are integrated to provide a volume. The program computes other measurements in a similar way. A complete list of measurements provided as expressed in conventional pulmonary function abbreviations and terminology is given in Table I.

Table I
Spirometry Measurements

Forced Vital Capacity in liters	FVC
Forced Expiratory Volume in one second in liters	FEV_1
Peak Expiratory Flow Rate in liters/second	PEFR
Maximum Midexpiratory Flow Rate in liters/second	MMFR
Forced Vital Capacity Time in seconds	FVCT
Inspiratory Capacity in liters	ICAP
Flow at 50% expiratory volume remaining in liters/second	$\dot{V}50$
Flow at 25% expiratory volume remaining in liters/second	$\dot{V}25$
$(\dot{V}25 - \dot{V}50)$ / FVC/4	
Rise-time between 10% and 90% total expiration in seconds	

The technician can display this list of measurements on the CRT under teletype control. If the values indicate that the subject did not make his maximal effort, he is again encouraged to do better. The forced vital capacity measurement is usually the one of greatest interest.

The above sequence of: 1) maneuver by subject, 2) technician inspection of the computer pattern recognition results, and 3) technician inspection of the computed measurements is repeated until the technician is satisfied with the results. He then instructs the computer to save the digitized flow data and derived

measurements on LINC tape for subsequent playback or further processing.

Other displays which can be requested by the technician under teletype control include the volume curve (i.e., the integrated flow curve) and a curve produced by plotting expiratory flow against volume. Normally the technician also requests a teletype listing of the subject identification information along with the derived measurements for a permanent hard copy record.

THE NEED FOR REMOTE SERVICE

The system described above served the lung function project well for two years. It became increasingly apparent, however, that the PDP-12 computer could be more advantageously applied to a diversity of biomedical projects that were being conducted at another site several miles away from the school. The question was how to continue to support the important growth of lung function project remotely.

ALTERNATIVES CONSIDERED

The alternative of transporting the test subjects from the school was rejected primarily because it would have greatly increased the time required for testing.

The alternative of providing the school with its own dedicated Central Processing Unit (CPU) was rejected on the basis of cost considering the following facts:

- A minimally configured PDP-12 costs $37,000
- DEC's PDP-8 based PTS system costs $25,000
- A PDP-11 or other non-DEC equipment would cost approximately $20,000 for hardware and would create prohibitive reprogramming costs.

ALTERNATIVE CHOSEN

The final choice was to provide the pulmonary function laboratory in the school with a storage-tube graphics terminal and a telecommunications link between the terminal and the PDP-12. Another telecommunications link transmits the analog signal from the spirometer amplifier to the PDP-12's A/D converter. The remote support configuration is shown in Figure 2. The purchase costs of the Tektronix 4010 graphics terminal and the DEC DP-12 interface totaled around $5,400.

The 603 analog modems are adequate for purposes of the project. They can transmit signals within a voltage range of ±2 volts and within a frequency range of 0-100 Hertz. Some noise problems were experienced, but they were solved satisfactorily by adding a simple low-pass filter circuit in the line going to the A/D converter.

The 202R modems operate at 1200 baud. This rate allows the display of the spirometer flow signal in real time at the Tektronix terminal.

Functionally, the remote system behaves exactly like the onsite system. The only noticeable difference is that several scans of flow data are displayed before erasing the screen. When a "processed" curve is displayed, it takes 3 or 4 seconds to be painted rather than the few milliseconds required on the original refresh display. The technicians have had no trouble in switching systems; in fact, when they have had a choice of systems, they have preferred the new one since it does not involve the noisy teletype, and the keyboard and display screen are positioned in the same field of vision.

One desirable feature of the system is that all displays appearing on the Tektronix screen also appear on the display of the home based PDP-12. This provides an effective means of allowing the medical staff at the home site to view what is happening at the school and to offer assistance over the telephone if required. It is also convenient for demonstration purposes.

FUTURE PLANS

Currently statistical analysis of the data is performed off-line on the Univac 1110. The data is communicated to the Univac in the form of punched cards prepared from forms hand transcribed from the PDP-12 teletype listings. The project team is considering bypassing this transcription step by copying the LINC tapes to 9-track magnetic tape, which can then be read directly by the Univac.

One of the highest priority future plans is to replace the very high-priced ($83/month total for the four modems) Bell-leased equipment with purchased equipment from an independent manufacturer. Bell was selected initially so as to reduce the number of vendors to be dealt with in the event of system problems. This conservative philosophy proved successful since it was possible to solve all the problems that did arise.

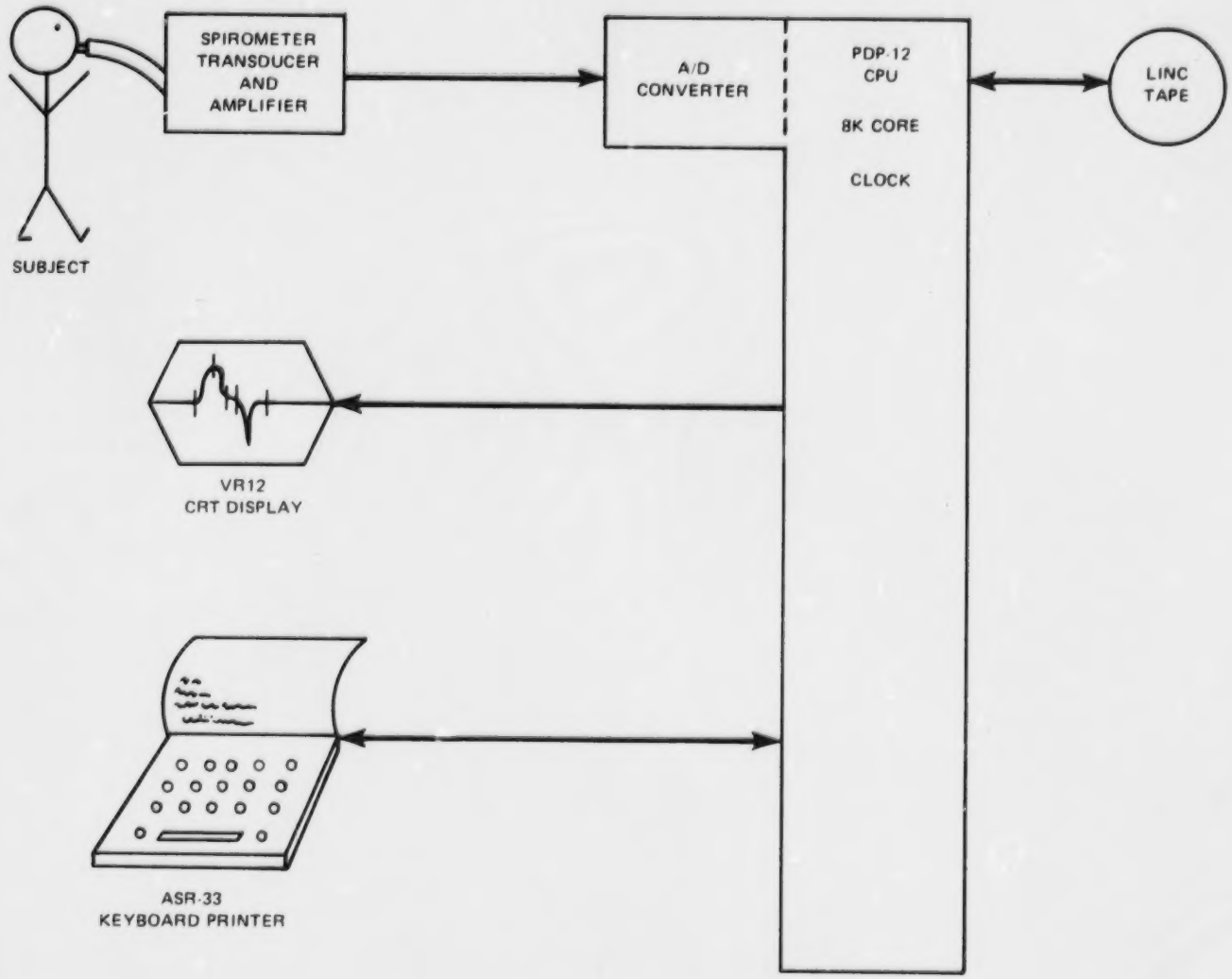


Figure 1
Onsite Configuration

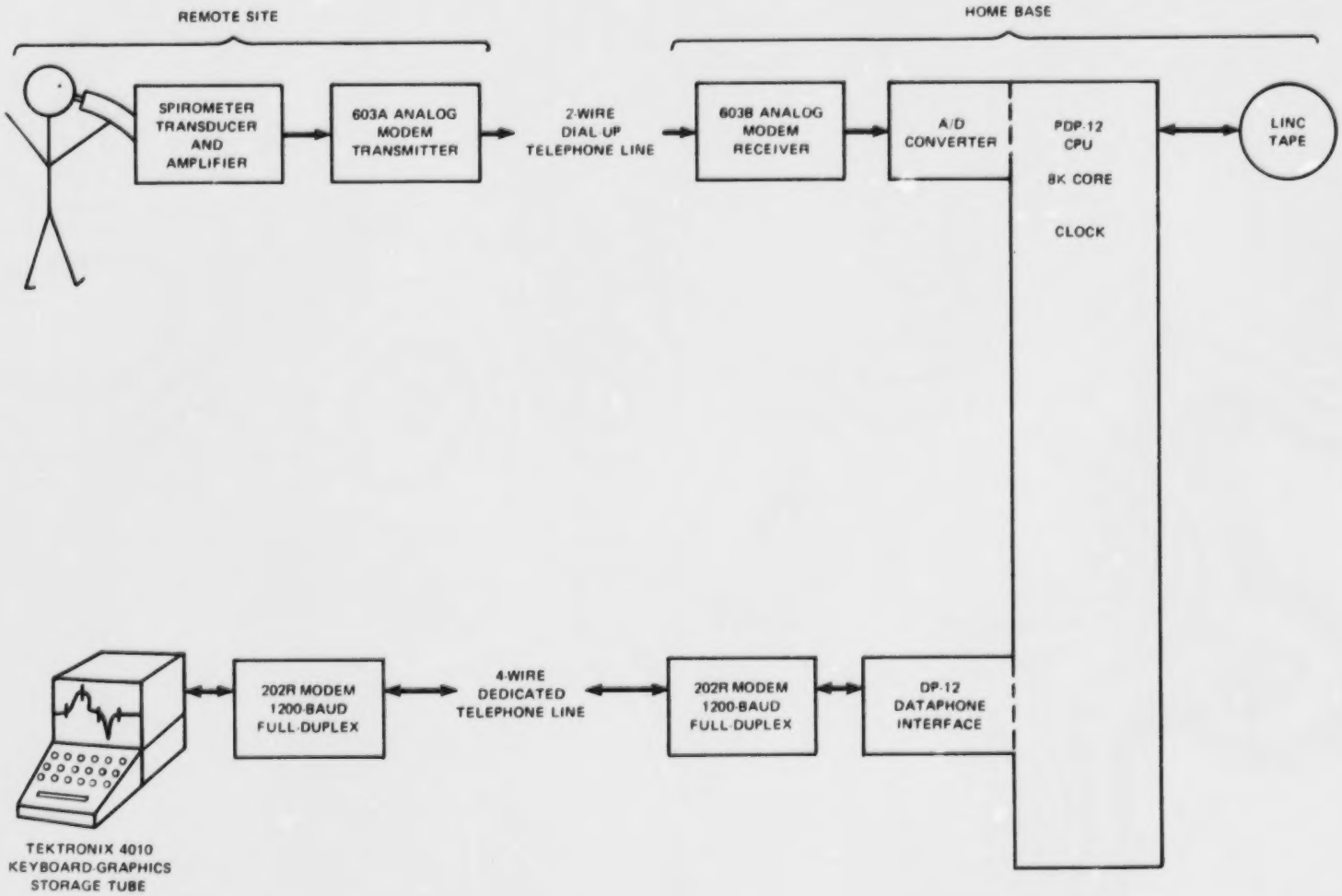


Figure 2
Remote Site Configuration

AN ON-LINE REAL-TIME MULTI-USER LABORATORY AUTOMATION SYSTEM

By William L. Budde,* Edward J. Nime, and Jack Teuschler

INTRODUCTION

Before 1965 it was unthinkable to put an electronic digital computer in an analytical chemistry laboratory. Since that time the steady decreases in the cost of computers and the increases in their reliability have brought about another revolution in analytical chemistry. Digital computers that were once in the $10,000 to $1,000,000 class now sell for $500 to $100,000 and have far more computational power. The analytical chemical instrumentation that has all but completely replaced the buret and filter paper is rapidly becoming computerized instrumentation.

Why is the digital computer invading the laboratory? As our technology and society have become more complex, the demand for more chemical analyses in many fields has increased exponentially. Along with this increased demand there are the requirements for more accuracy, better precision, higher sensitivity, more timely results, greater selectivity, and of course all of these at a lower cost per analysis. A good example of this is the health field. At one time the family physician's stethoscope was one of the few routinely used diagnostic tools. Today a large clinical lab must do literally thousands of blood cholesterol and urine sugar analyses every day.

In the environmental field the measurement of specific air and water pollutant chemicals is the basis for the whole environmental movement. Until reliable measurements were made and correlated with undesirable health or wildlife population effects, environmental concern was mostly limited to those concerned with purely aesthetic values. Currently there is considerable emphasis on setting standards for acceptable air and water quality, issuing permits for discharge of wastes into rivers and oceans, monitoring these effluents to ensure compliance with permit limitations, and conducting enforcement actions when violations occur. All of these activities are increasing the demand for more and better chemical environmental analyses.

Better analyses embody the ideas of accuracy and precision, and require extensive use of analytical quality control techniques. Quality control is often deleted in analytical laboratories because of its cost and time requirement. With this deletion, the meaningfulness of the measurements decreases substantially. There is nothing more worthless than the wrong answer. With a computerized system, quality control is easily possible without a high price. Another aspect of better analyses is the desire for new kinds of measurements that are more revealing about the state of environmental pollution than traditional measurements. These more revealing measurements are often more complex and simply cannot be accomplished economically or at all without on-line computerization.

In response to the needs of the Environmental Protection Agency (EPA), the Methods Development and Quality Assurance Research Laboratory (MDQARL) of the National Environmental Research Center (NERC) in Cincinnati and the Computer Services and Systems Division of NERC-Cincinnati began a program to satisfy these needs. The primary mission of MDQARL is to develop, improve, and validate methodology for the collection of physical, chemical, radiological, microbiological, and biological water quality data by EPA Regional, Office of Enforcement and General Counsel, Office of Air and Water Programs, and other NERC-Cincinnati organizations. This laboratory also advises EPA and other laboratories concerning the development of monitoring and quality control programs for water quality activities. These functions in methodology development and quality control are clearly highly relevant to the development of a laboratory automation system.

PROJECT GOALS

In late 1972 an interagency agreement was established with the Lawrence Livermore Laboratory (LLL) of the Atomic Energy Commission (AEC) which is operated for AEC by the University of California. This organization was chosen to assist in the project because of its long record of solid accomplishments in this field. Some of the earliest successful applications of computers in a chemical laboratory were developed in this organization's general chemistry division.† A thorough systems analysis was performed to define precisely the needs of

*Speaker

†See J.W. Frazer, *Science and Technology*, July 1968, p.41. Also see J.W. Frazer, *Anal. Chem.*, *40* 26 A (1968).

EPA, to establish the exact goals of the project, to write detailed specifications for hardware and software, and to develop an implementation plan. Several goals that were defined are:

1. To develop a laboratory automation system that would incorporate presently owned chemical analysis instrumentation widely used at MDQARL and many other EPA laboratories for measurements of water quality parameters.

2. To develop this methodology in a fashion that will permit the adaptation of the technology to other EPA laboratories at a very significant cost and time savings. In particular, the designs for hardware interfaces between the instruments and the computer and the custom software would become EPA property and could be used in any EPA laboratory without further cost.

3. To develop an open ended design that would permit the attachment of many additional instruments that might be used for a variety of measurements including nonwater parameters.

4. To write as many of the computer programs as technically feasible in a very flexible high-level modern programming language. This would support the ease of modification and improvement of the software by scientific personnel and the transfer of technology to other laboratories.

5. To design the system with sufficient flexibility so it is applicable to methods development research as well as the production atmosphere. In a methods research program, automation makes careful testing of new procedures possible by allowing independent variation of a large number of method variables with a statistically significant number of samples.

INSTRUMENTATION TO BE AUTOMATED, PHASE I

In the first phase of the MDQARL pilot project, five instruments were selected for automation. These were a Perkin-Elmer model 503 atomic absorption spectrometer, a Varian model AA-5 atomic absorption spectrometer, a Jarrell-Ash 3.4 meter electronic readout emission spectrometer, a Technicon AutoAnalyzer II, and a Beckman total organic carbon analyzer.

The first two instruments were chosen because atomic absorption (AA) spectrometry is the most widely used method for measuring trace metal concentrations. The two manufacturers represent a high percentage of the total market for this equipment. EPA laboratories utilize a large number of AA spectrometers with few, if any, automated for on-line data acquisition.

Although EPA uses only a few emission spectrometers for trace metal analyses, this instrument was included because it has simultaneous, multielement capability and should be a method of choice in monitoring for many elements when large numbers of samples are involved. Emission spectrometry will be used more in the future as newer and more sensitive techniques are developed. The Technicon AutoAnalyzer is widely used for measurements of many pollutants including the important nutrients.* The total organic carbon instrument is widely used to measure the organic load in many waste streams.

Atomic Absorption Spectrometers

These units measure trace metal concentrations by optical spectrometry using samples obtained by one of two methods. Water samples may be aspirated into a flame or gaseous samples may be swept from a graphite furnace. Water samples can be retrieved from a computer controlled 40 position circular sample vial holder. Sampling modes available with this unit include semi-automatic sampling where the operator moves the sample holder by entering commands from his terminal, fully automatic sampling where the computer automatically moves the sample holder, and fully automatic bracketing of the unknown with appropriate standards. The computer program that reads the data is an assembly language subroutine that is called from the high-level BASIC language. Data reduction to concentration can be accomplished, at the user's option, by interpolation or by first or second degree linear least squares fitting. The operator can enter a wide range of information about his samples from his terminal. This information can also be output along with the analytical results in a variety of formats to suit the particular situation.

Emission Spectrometer

The automation of this instrument is somewhat different from the other four instruments since the presently existing analog to digital converter that drives a nixie tube digital display will be utilized. In the digital display system there is a high degree of operator-

*Methods for Chemical Analyses of Water and Wastes, Environmental Protection Agency, 1974.

instrument interaction. The operator manually increments a sequential readout from the 23 data channels, copies digits from the display onto a raw data report form, conducts calculations, makes judgments about the result, and begins a new sequence using the results of the judgments. This very desirable interactive feature is retained in the automated system, but the operator is released from the manual incrementing, reading, copying, and calculating functions. He retains judgment functions based on reduced data that is dynamically displayed at the instrument terminal. The automated system greatly improves instrument calibration, background correction, quality control, and development of analysis reports. As in the other instruments, information about the samples may be entered from the operator's terminal keyboard and included in part or in whole on the analysis report.

Technicon Autoanalyzer

This system results in a digital computer readout of the Technicon AutoAnalyzer output as it aspirates standards and samples from the existing automatic sampler module. The computer measures the time between several standard peak heights and, using this time, looks for absorption peaks above background at appropriate intervals as samples pass through the colorimeter. A slope measurement is made to ascertain that the readings start on the upside of a peak. The maximum value of these readings is selected as the datum point to be associated with that particular sample.

Unknown concentrations are reported as the analysis proceeds. Unknowns out of maximum range of the instrument cause the computer to ring a bell and the operator may add a diluted sample at the end of the series. The computer looks for these samples automatically. Dilution factors are entered at the completion of all samples prior to obtaining a formal analysis report.

Provisions are made for quality control checks using standards and spiked samples. The pattern for quality control checks is predetermined and the cycle of repetition entered by the operator. Quality control chart values for acceptance are stored. The computer assigns automatic sampler position numbers on the basis of the number of standards and the sequence of the unknown identification information entered. Calculations of concentrations are done by first, second or third degree least squares fitting of the standard curve or by simple interpolation as predetermined by the operator. The type of fit may be selected at the beginning of the analysis or after standards have been read.

Total Organic Carbon Analyzer

This system results in a digital computer analysis readout of the Total Organic Carbon Analyzer data as samples or standards are injected into the furnace. The computer searches for a peak height or integrates under a curve as the CO_2 from the sample or standard is swept through the infrared detector. Replicate injections are averaged and after each injection, the new average is reported. Individual results may be replaced without losing prior information or, alternatively, an entire sample or standard run may be deleted.

Allowances are made for the introduction of standards as checks and also for samples spiked with known amounts. Quality control chart values are stored. Calculations of concentrations are done by first or second degree least squares fitting of the standards data or by simple interpolation using the two closest standards as predetermined by the operator.

TYPE OF COMPUTER SYSTEM

A fundamental decision that had to be made was the choice between a system of dedicated minicomputers or microcomputers, an in-laboratory time sharing computer, or a remote time sharing system. The remote time sharing concept was easily eliminated as this was extensively tested during the late 1960's and was proved a failure. In laboratory systems, a central processing unit (CPU) response time of the order of microseconds to milliseconds is absolutely essential to avoid loss of data from instruments. Response times of this speed are simply well beyond current capabilities. In addition the difficulties and cost of transmission of analog data and the control of the computer by individuals not familiar with laboratory operations led to a totally unacceptable level of reliability.

A system of minicomputers or microcomputers was eliminated because of the cost of equipping each with sufficient memory and peripheral devices. Also small systems invariably must be programmed in assembly language which precludes fast modification and updating of programs. Specifications for a medium scale in-laboratory time sharing system are developed and are summarized below.

Computer Hardware Specifications Summary

A 12-18 bit word length processor equipped with floating point hardware, integer multiply and divide hardware, and with capabilities necessary to interface a

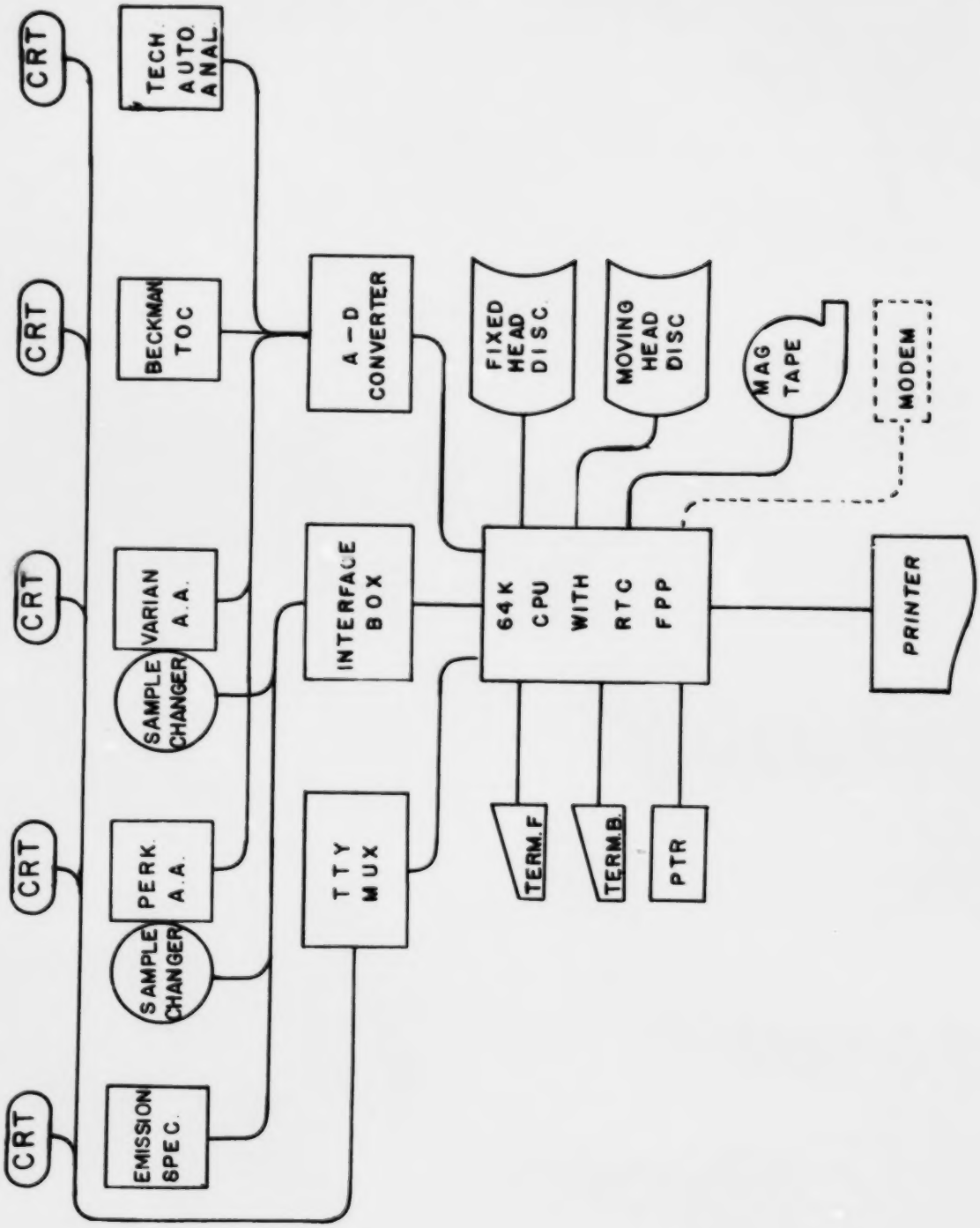
fixed head high-speed swapping disk, a moving head cartridge disk, industry standard magnetic tape, a fast paper tape reader, a medium-speed line printer, a console keyboard-printer, remote cathode ray tube (CRT) terminals, and multiplexed A/D and D/A converters.

Computer Software Specifications Summary

A well-documented and field tested operating system which, in a time slicing sense, can serve up to ten terminal users simultaneously. The operating system must support an extended version of the standard Dartmouth BASIC interpreter language. The BASIC language must be capable of acquiring data in real time, controlling instruments in real time, processing character strings, manipulating data files, supporting insertion of custom input-output device handlers, supporting insertion of machine language subroutine calls, chaining to other programs, and spooling output.

In an open competitive bidding process, the Data General Corporation was the successful bidder with a NOVA model 840 computer and associated peripherals. The hardware configuration is shown in Figure 1. Figure 2 gives a general overview of the organization of the laboratory user software. Figure 3 is the organization of the Data General real-time disk operating system which meets all specifications. Figure 4 gives a more detailed outline of the foreground core memory where the BASIC language and assembly language instrument handlers reside. Figures 5 and 6 show the significant features of BASIC and a comparison with FORTRAN. All of these were discussed in greater detail during the presentation and subsequent sections of this report.

Figure 1
Hardware Configuration for the NERC-Cincinnati MDQARL Laboratory Automation Pilot System



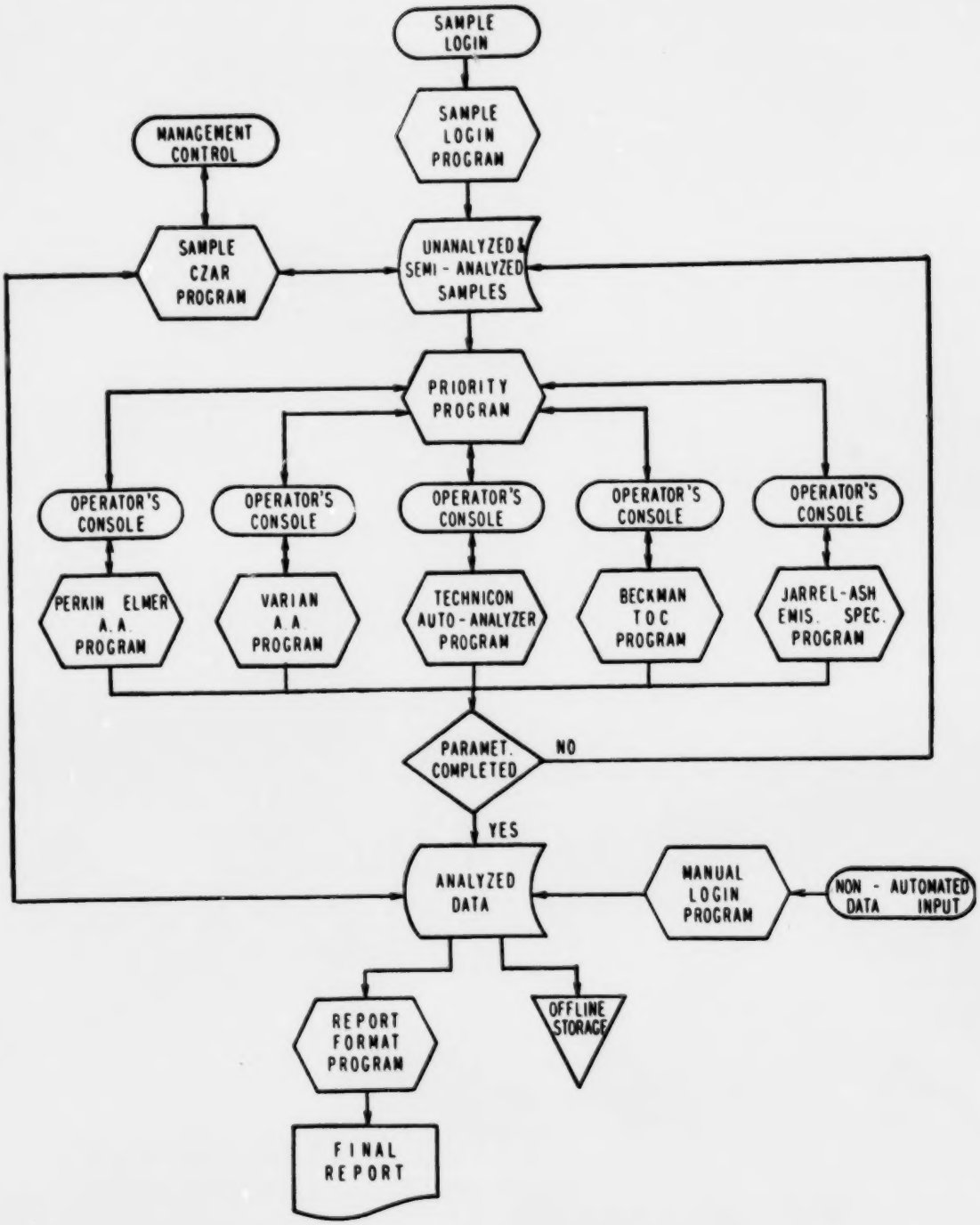


Figure 2
Software Organization for the NERC-Cincinnati MDQARL Laboratory Automation Pilot System

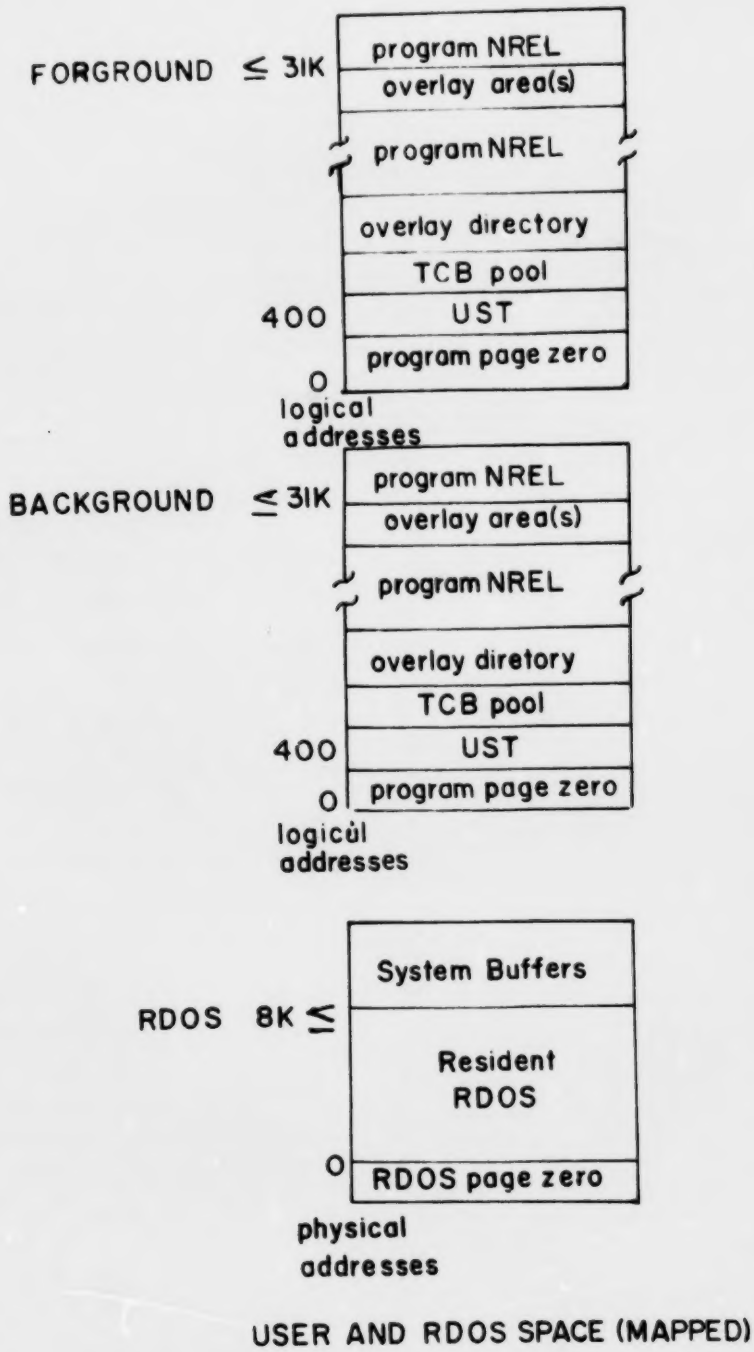


Figure 3
Organization of Core Memory in the Data General Real Time Disk Operating System (RDOS) with Hardware Memory Management

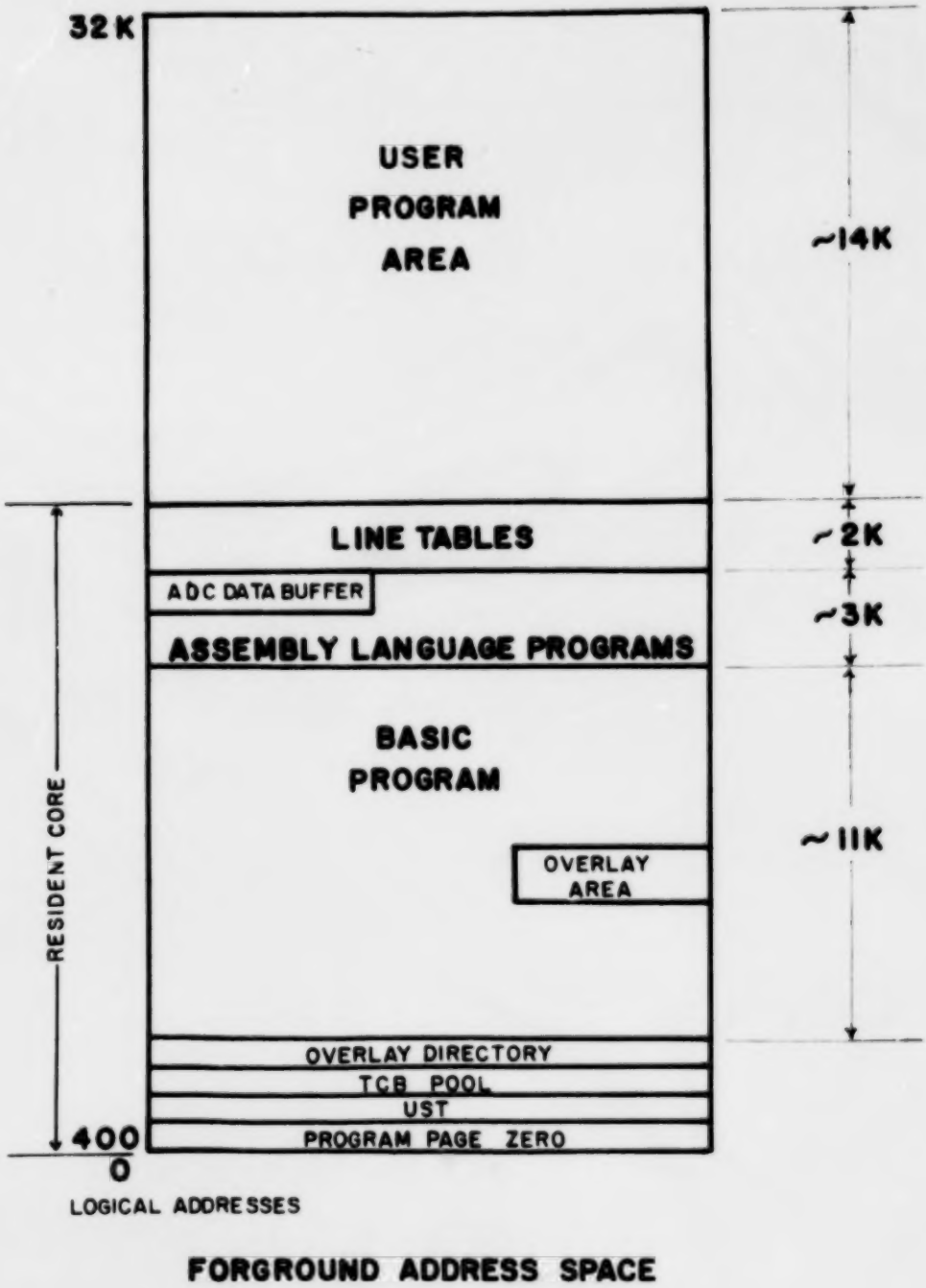


Figure 4
Organization of Foreground Core Memory Under RDOS in the NERC-Cincinnati MDQARL Laboratory Automation Pilot System

FEATURES OF EXTENDED BASIC

1) MATRIX MANIPULATION COMMANDS

2) STRING COMMANDS

3) FILE I/O HANDLING

4) ASSEMBLY LANGUAGE CALLS FROM THE BASIC LANGUAGE

5) CHAINING CAPABILITY

6) TIME SHARING CAPABILITY

Figure 5
Significant Features of Data General
Multi-Tasking Extended BASIC Programming Language

PROGRAMMING LANGUAGES

BASIC

```
FIND SOURCE FILE
CALL EDITOR
INSERT : I2 = A*B

INSERT: PRINT "THE ANSWER IS";I2

END EDIT

RUN
```

FORTRAN

```
    FIND SOURCE FILE
    CALL EDITOR
    INSERT: ISTREAK = A*B
    REAL ISTREAK
    INSERT: WRITE (3,4) ISTREAK
4   FORMAT (15H THE ANSWER IS ,F5.2)
    END EDIT
    REQUEST COMPILATION
    RUN COMPILED CODE
```

Figure 6
Comparison of Programmer Unit Operations in Modifying FORTRAN and BASIC Source Code

COMPUTERIZED CHROMATOGRAPHY-MASS SPECTROMETRY

By William L. Budde and D. Craig Shew*

BACKGROUND

Identifying and measuring specific organic compounds that contaminate the environment has concerned environmental research scientists for many years. The huge number of potentially harmful or toxic compounds that are theoretically possible has discouraged the development of a specific analytical method for each. Instead, general approaches based on slight differences in solubility, chromatographic behavior, and spectroscopic measurements were developed and used widely to gain information about the types of compounds present in various samples.

Unfortunately, many of these standard techniques of chemistry, which provided very valuable information in flagrant pollution situations, were either slow, expensive in terms of skilled manpower, or relatively insensitive to very small quantities. In addition, most of these techniques frequently failed to generate really definitive information. As a result, environmental researchers began to turn to a new multidisciplinary technique: computerized (automated data processing) gas chromatography/mass spectrometry (ADP/GC/MS).

This sophisticated technique was developed during the 1960's primarily to meet the demand of basic research workers in medical, biochemical, flavor, odor, medicinal, and geochemical research. Basic research supported by the National Institutes of Health, the National Science Foundation, and the National Aeronautics and Space Administration played the major role in the development of the concept. Since one of the basic problems in each of the above areas of research is the identification and measurement of trace quantities of a great variety of organic compounds, the application of the technique to environmental research was obvious but slow to come about.

Like any automated method, ADP/GC/MS does not reduce the need for skilled manpower; in fact, it requires highly skilled, multidisciplinary teams of spectroscopists, analytical chemists, electronics engineers, and laboratory minicomputer specialists. The major advantage of this method is that it enormously increases the capacity of a fixed or limited staff to make large numbers of unambiguous identifications of specific organic compounds in environmental samples. Identification of noxious pollutants at the part-per-billion level with a high degree of confidence in the result has become nearly routine in more than a few U.S. Environmental Protection Agency (EPA) laboratories. What was once a nearly impossible task for a staff of 100 working for 6 months can sometimes be accomplished by several people in a few hours. The ADP/GC/MS system is by far one of the most outstanding examples of the successful integration of laboratory-scale, analytical instrumentation with a dedicated data processing system.

STATE OF THE ART

EPA has made a major commitment to ADP/GC/MS during the last few years by installing about 23 complete systems valued at about $2.5 million in laboratories across the country. The systems are employed in eight regional surveillance and analysis laboratories, both National Field Investigations Centers, two National Environmental Research Center (NERC) - Cincinnati laboratories, one NERC-RTP laboratory, six NERC-Corvallis laboratories, and two Office of Pesticide Programs laboratories. The annual total costs dedicated to these systems should be at least equivalent to the purchase price.

Notwithstanding the many different missions of the various EPA laboratories across the country, ADP/GC/MS has proved to be a highly successful technique capable of accomplishing a wide variety of tasks. In many cases involving the identification of organic pollutants, ADP/GC/MS is not only the method of choice, it is, indeed, the only method available. Typical of the wide variety of uses are the following:

- Identification of organic compounds in industrial effluents for background data and/or enforcement action
- Characterization of oil spills
- Identification of organic compounds as the cause of taste and odor in drinking water supplies

*Speaker

- Evaluation of the effectiveness of treatment systems designed to remove organic pollutants
- Studying the accumulation of organic pollutants in aquatic life
- Determining the causes of fish kills.

SYSTEM OPERATION

Eighteen of the systems mentioned above are nearly identical Finnigan quadrupole mass spectrometers* computerized with Digital Equipment Corporation PDP-8 minicomputers (Figure 1). The three functional parts of the system are as follows:

1. The gas chromatograph is very well-established as a powerful tool for the separation of complex mixtures of volatile organic compounds, while the sample enrichment device provides an effective interface between the gas chromatograph and mass spectrometer.

2. The mass spectrometer is the only GC detector used, and like conventional GC detectors, it is very sensitive. In contrast to the single-channel response of conventional GC detectors, the mass spectrometer provides a multichannel response that carries a great deal of information about the molecular structure and composition of organic compounds. This information may be displayed graphically by a fast cathode-ray tube or a slow (pen and ink) plotter, printed in digital form, or transmitted over conventional voice-grade telephone circuits to other data-handling systems.

3. The minicomputer controls the operation of the mass spectrometer and processes the generated data. In noncomputerized systems, mass spectra generally are not acquired continuously during the GC run. However, with a computerized system, the minicomputer simultaneously controls the operation of the mass spectrometer and continuously acquires mass spectral data. This large volume of MS data is processed by the computer and stored temporarily on a magnetic disk or magnetic tape before plotting, printing, or transmission. Personnel who previously devoted long, unproductive hours to manual data processing are freed to do productive tasks such as data interpretation.

Interpretation of Data

The necessary prerequisite for effective data interpretation in ADP/GC/MS is the availability of valid mass spectra. Assuming that the ADP/GC/MS is well-tuned and in good operating condition, the only other requirement for valid mass spectra is the clean separation of the individual components of the mixture on the GC column. This is, of course, the classical problem of chromatography, but it is made somewhat more manageable with the MS detector. An experienced user can frequently ascertain that the separation is clean and free of overlaps by an examination of the internal consistency of the mass spectral data. This is not possible with the simpler, conventional GC detector.

There are two fundamentally different approaches to the identification of specific organic pollutants from valid mass spectra. Interpretations may be attempted based on the theory of mass spectroscopy and the rules of fragmentation of organic ions in the gas phase. Unfortunately, this approach is frequently slow and tedious and requires extensive training and experience. However, it is generally far more definitive than the equivalent approach with other types of spectrometry (e.g., infrared spectra). This approach is clearly justified when all else fails in important situations such as an enforcement action or fish kill.

The alternative is the purely empirical method of searching a file of reference mass spectra to find a similar or exact match of the experimental spectrum. While manual searching of printed data is rather slow, subject to human error, and intellectually unstimulating, computerized searching overcomes these problems. Two experimental remote systems, which are markedly different in their approach to the problem, are currently in use by EPA and are described briefly below. Perhaps the simplest approach would be to utilize the PDP-8 minicomputer with the disk or tape storage available on the MS data system to search a data base of reference spectra. This may be the most economical method, but it has not been fully developed and is not currently used by EPA.

One system used by EPA was developed by the Battelle Memorial Institute with funding from the Southeast Environmental Research Laboratory component of NERC-Corvallis. It is intended to be used by a laboratory technician who needs to know the correct buttons to operate and the correct commands to enter. MS data are transmitted in batches directly from the ADP/GC/MS data system to a large Control Data Corporation model 6400 computer at Battelle-Columbus via standard voice-grade telephone lines. The system is highly automated, rather sophisticated, and makes good

*Mention of commercial products does not imply endorsement by the EPA.

use of the data. It is simple to use because it either finds a match or it does not; however, if no match is found, there are no other alternative modes of search. Since the system does require the ADP/GC/MS data system, use of the ADP/GC/MS is precluded during the matching operation.

The other system used by EPA was developed by the National Institutes of Health (NIH) and may be used by any EPA laboratory under an existing EPA-NIH interagency agreement. This system runs on a large Digital Equipment Corporation model PDP-10 computer at the NIH computer center in Bethesda, Maryland. The user enters MS data via a standard keyboard/printer terminal connected to a standard voice-grade telephone line and does not utilize the resources of the ADP/GC/MS data system. The matching system is highly interactive and conversational, and the user must impose his judgment in entering data. This system is oriented to the experienced user who sits in the interactive loop with the computer program and has many optional approaches to probing the data base and searching for a match. This approach makes less efficient use of computer resources and depends more on an operator's judgment to make good use of the available data. However, it is a more flexible method, especially when an exact match is not available. For economy and additional flexibility, these two experimental systems are in the process of being combined to use the same data base on the same time-sharing computer.

EPA-MS USERS GROUP

In the spring of 1972, an EPA-MS users group was organized for the purpose of promoting the informal exchange of technical information among the EPA laboratories using the 23 computerized GC/MS systems. The group is designed to be especially helpful to laboratory personnel (e.g., regional surveillance and analysis chemists) who are just beginning to acquire the skills necessary to utilize this sophisticated equipment effectively. In addition, the group keeps all involved personnel up-to-date on the latest instrumental improvements, new techniques, new equipment additions, and new software programs and capabilities.

CONCLUSION

It has been clearly demonstrated that ADP/GC/MS can provide extremely high quality and highly significant information about organic environmental pollutants to environmental scientists and enforcement officials. However, the achievement of this capability requires a substantial laboratory management commitment.

The ADP/GC/MS is not the kind of equipment that can stand idle for weeks and then be used effectively by occasional users. Users require in-depth knowledge, and the equipment requires routine maintenance, adjustment, and utilization. Furthermore, the commercial packages are not perfected and will require updating in the future. Nevertheless, it is believed that the commitment is worthwhile and that ADP/GC/MS routine analyses should replace, in many EPA laboratories, some of the traditional, environmental pollution methodology that generates less relevant and very much less significant data.

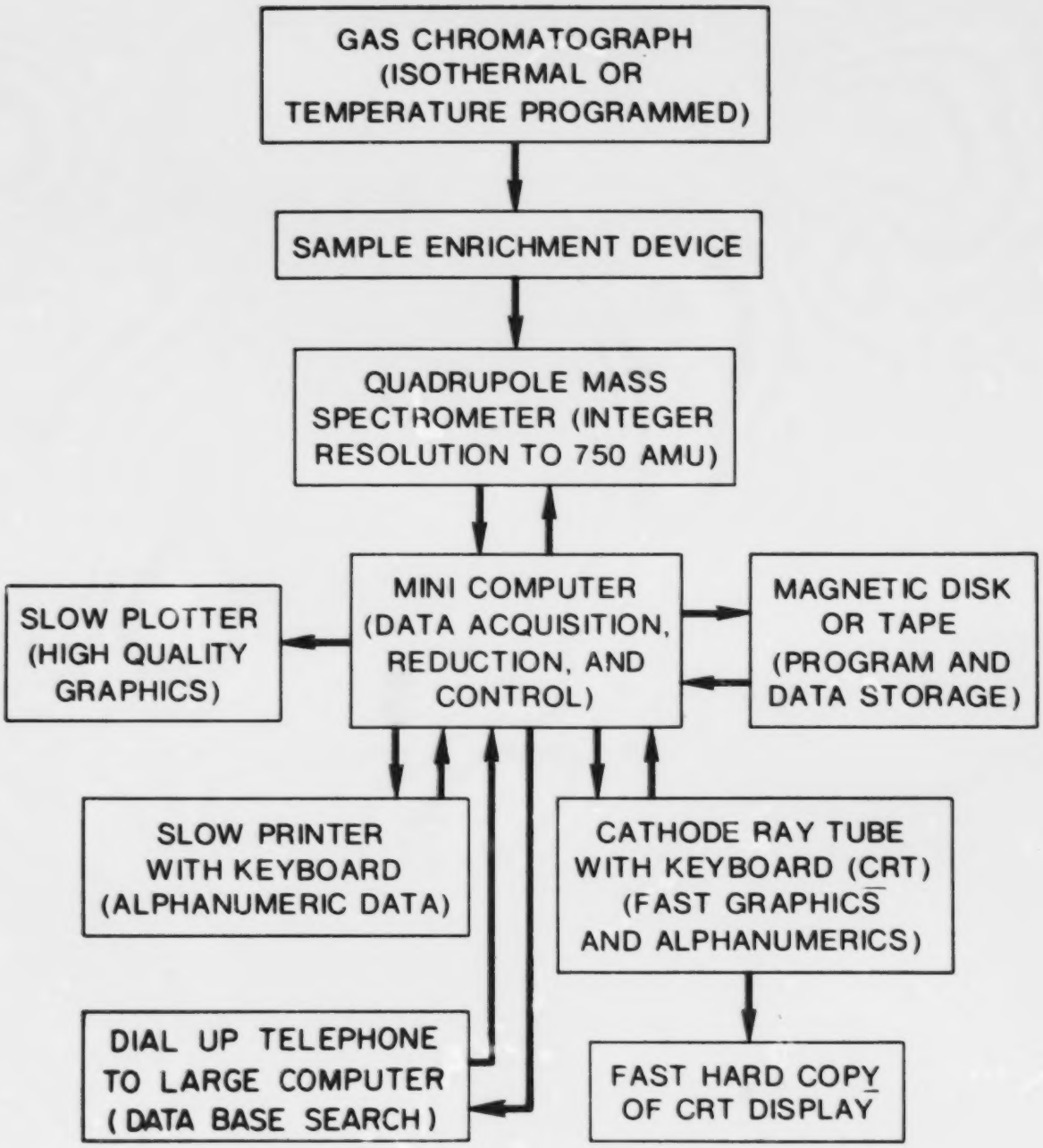


Figure 1
Configuration of the ADP/GC/MS System Widely Used in EPA

CALCOMP SOFTWARE

By Theodore R. Harris

CALCOMP software leased by the Environmental Protection Agency (EPA) was discussed, and its availability on the two major computer resources was portrayed. Utilizing FORTRAN, CALCOMP software provides the capability to produce graphic outputs. These outputs commonly are in the form of off-line pen plots.

CALCOMP provides three general classifications of graphic software. The highest level is made up of applications programs. These programs are as follows:

- GPCP - General Purpose Contouring Package (Figure 1)
- THREE D - A Perspective Three-Dimensional Package (Figure 2).

The second type of graphic software is classified as Functional Software. Functional programs or subprograms perform plotting functions frequently used in many different applications and are as follows:

- FLOCT - Flow Charting Package (Figure 3)
- FORGN - Form Design Package (Figure 4)
- DRAFT - Drafting Subroutine Package (Figure 5)
- SCIEN - Scientific Subroutine Package (Figures 6 and 7)
- GENRL - General Subroutine Package (Figures 8 and 9)
- BUSNS - Business Subroutine Package (Figure 10)
- DRIVER - Curve Fitting Subroutine Package (Figures 11 and 12).

The functional subroutines are shown in Figure 13.

The third classification is the Basic Software, which is the lowest level for controlling the plotting hardware. These routines are as follows:

- PLOT - Direct control of pen movement
- SYMBOL - Produces annotation
- NUMBER - Produces number conversion for annotation
- SCALE - Determines range and scaling factor of data
- AXIS - Produces X and Y axis elements
- LINE - Produces line plots.

The software is currently installed at Research Triangle Computing Center and is being installed at Optimum Systems Inc. (OSI). A training seminar will be announced in the "Data Systems Newsletter."

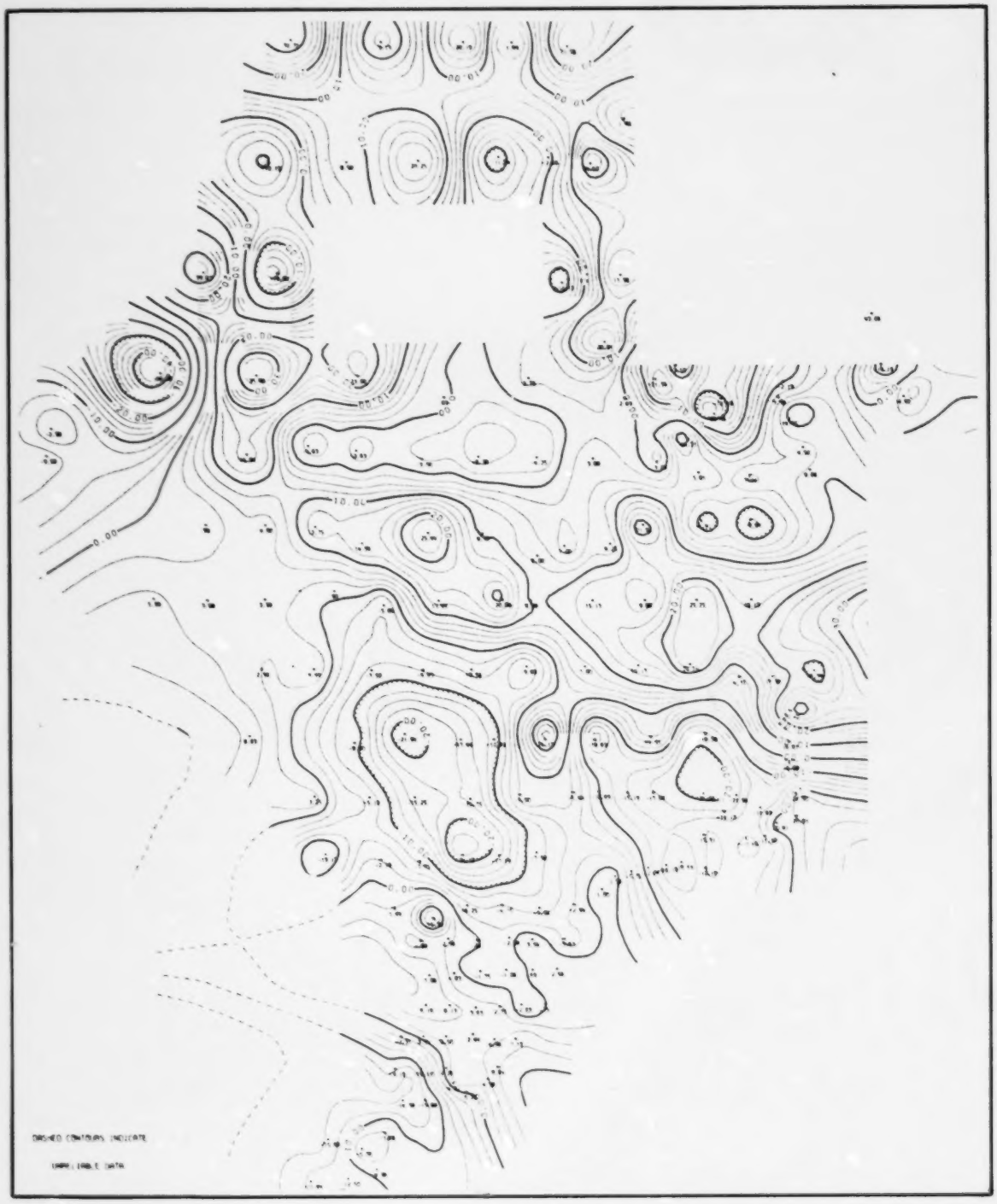


Figure 1
Flexible Mode Sample Problem Output

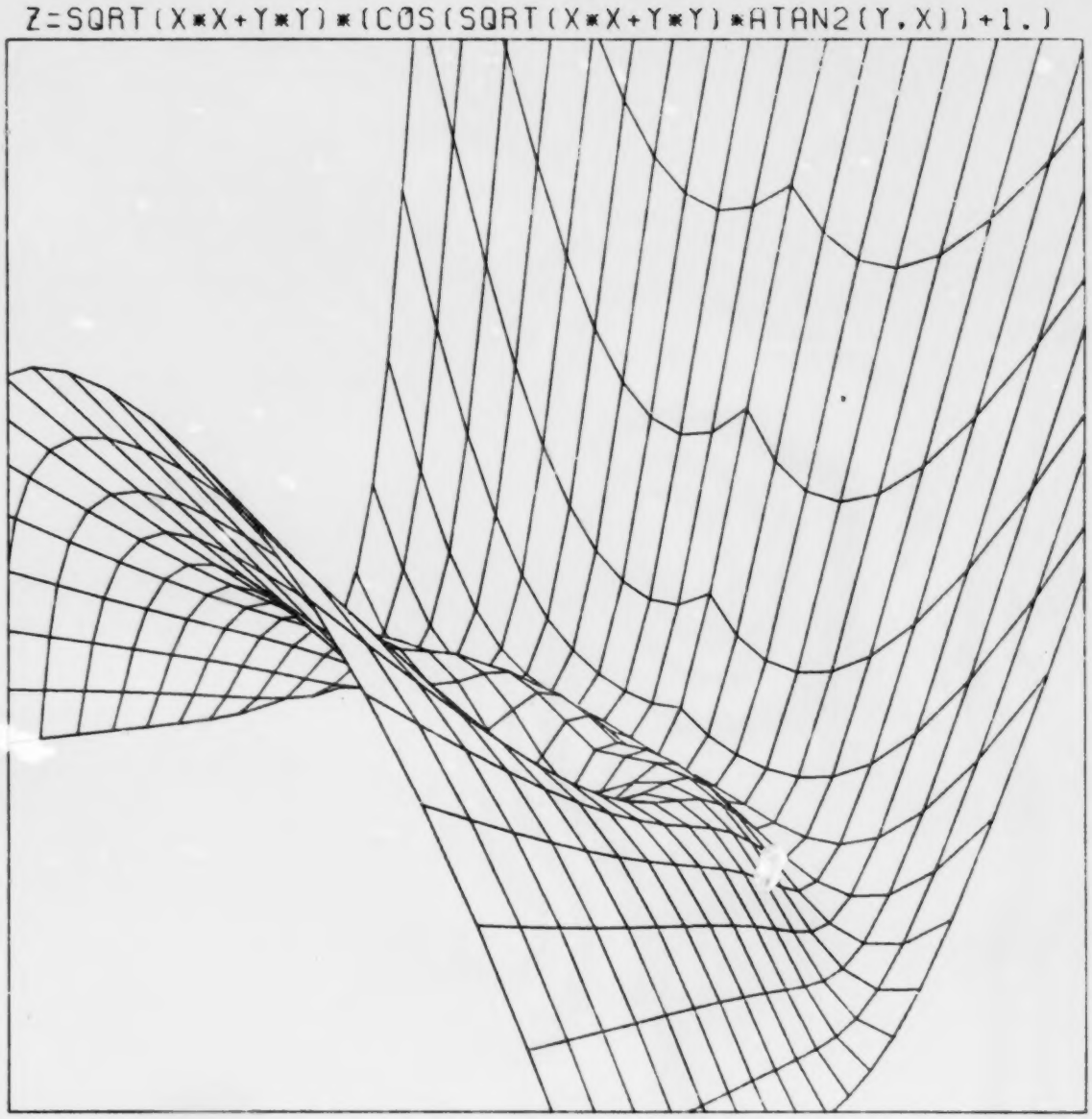


Figure 2
Sample Perspective View
Q = 2, THETA = 40°, GAMMA = 50°
Hidden Lines Eliminated

ILLUSTRATION OF FLOCT FACILITIES

```
SUPPLEMENTARY NT#9 CODES.
NCD #    1-988 ALLOWS THAT NUMBER OF FREE FORM
                COMMENT LINES.
NCD # 989      MOVES WITH PEN UP TO X,Y.
NCD # 990      PLOTS SOLID LINE FROM CURRENT
                POSITION TO X,Y.
NCD # 991-999  PLOTS .1-.9 DASHES RESPECTIVELY
                FROM CURRENT POSITION TO X,Y.
```

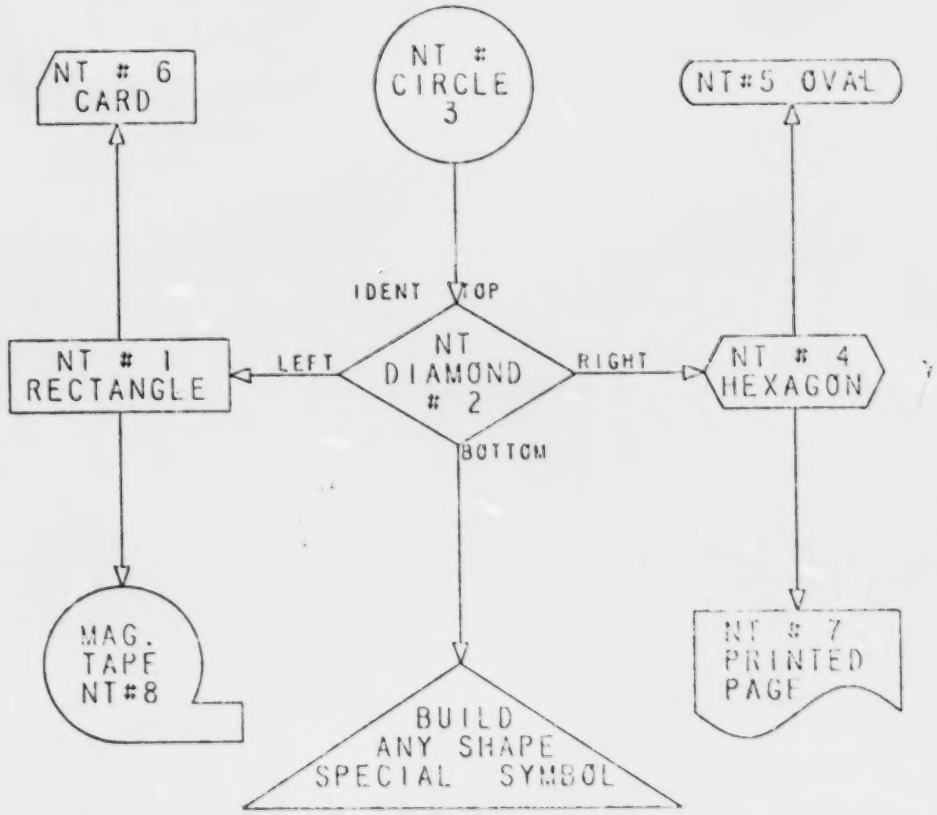


Figure 3
Illustration of FLOCT Facilities

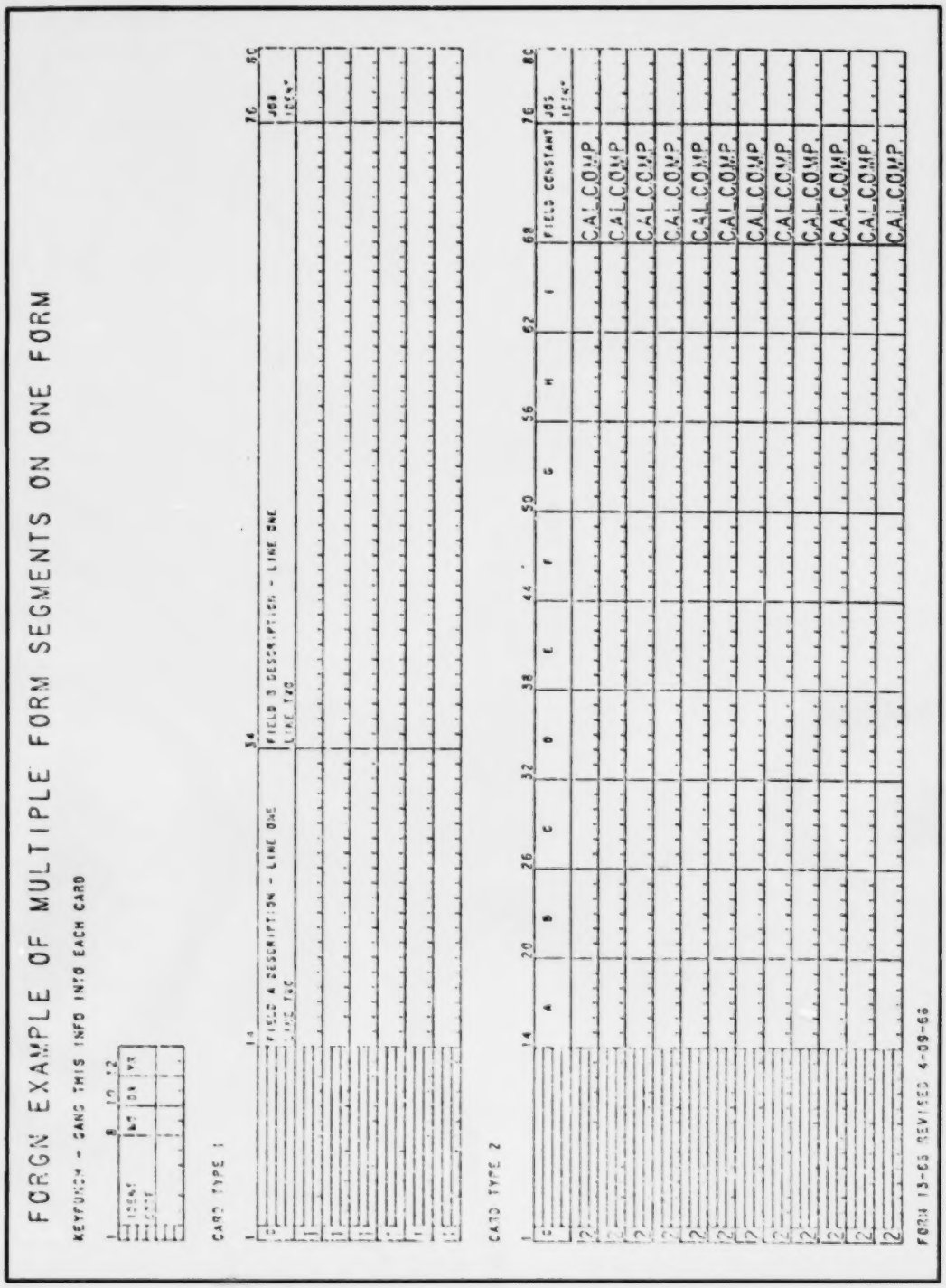


Figure 4
FORGN Example of Multiple Form Segments on One Form

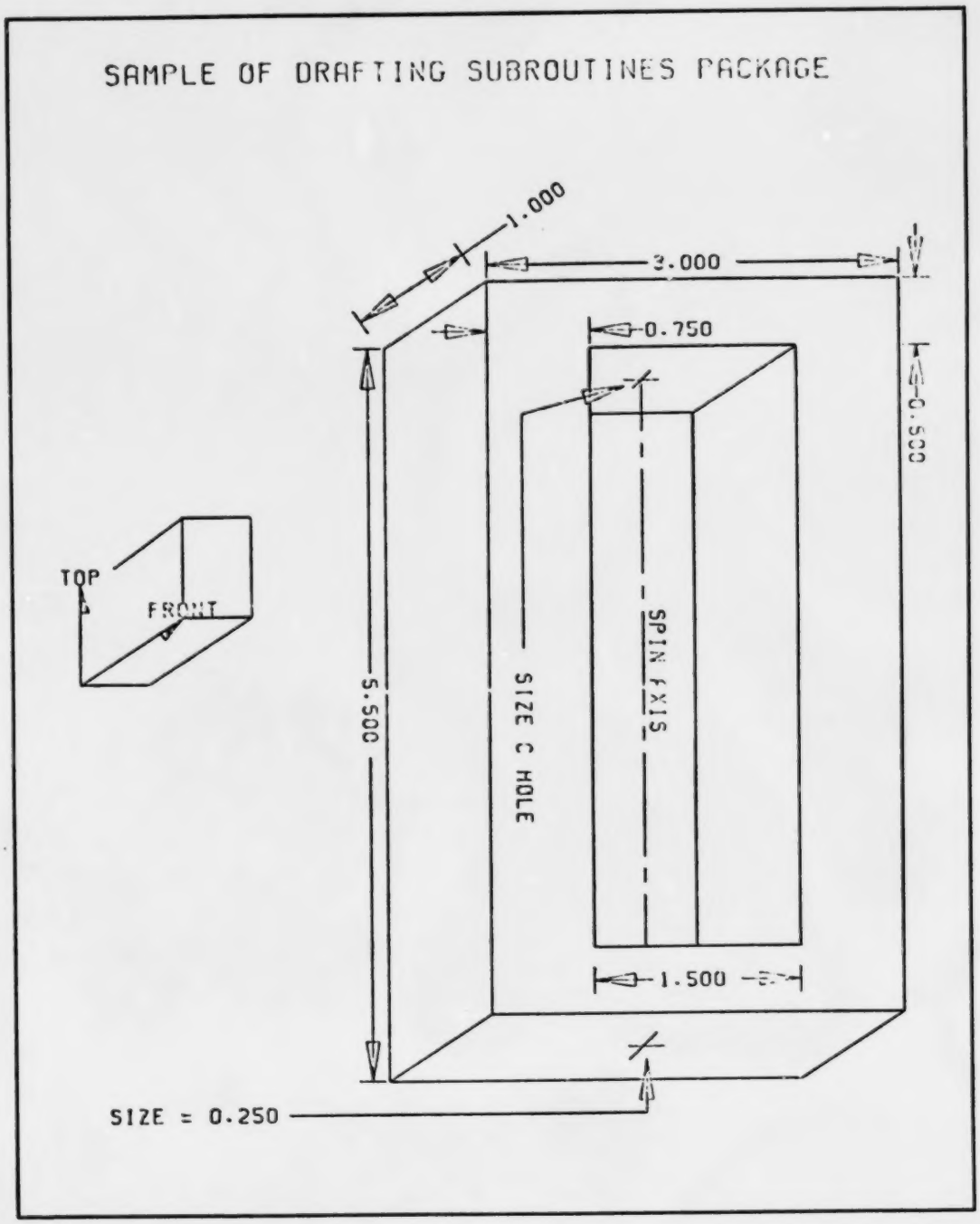


Figure 5
Sample of Drafting Subroutines Package

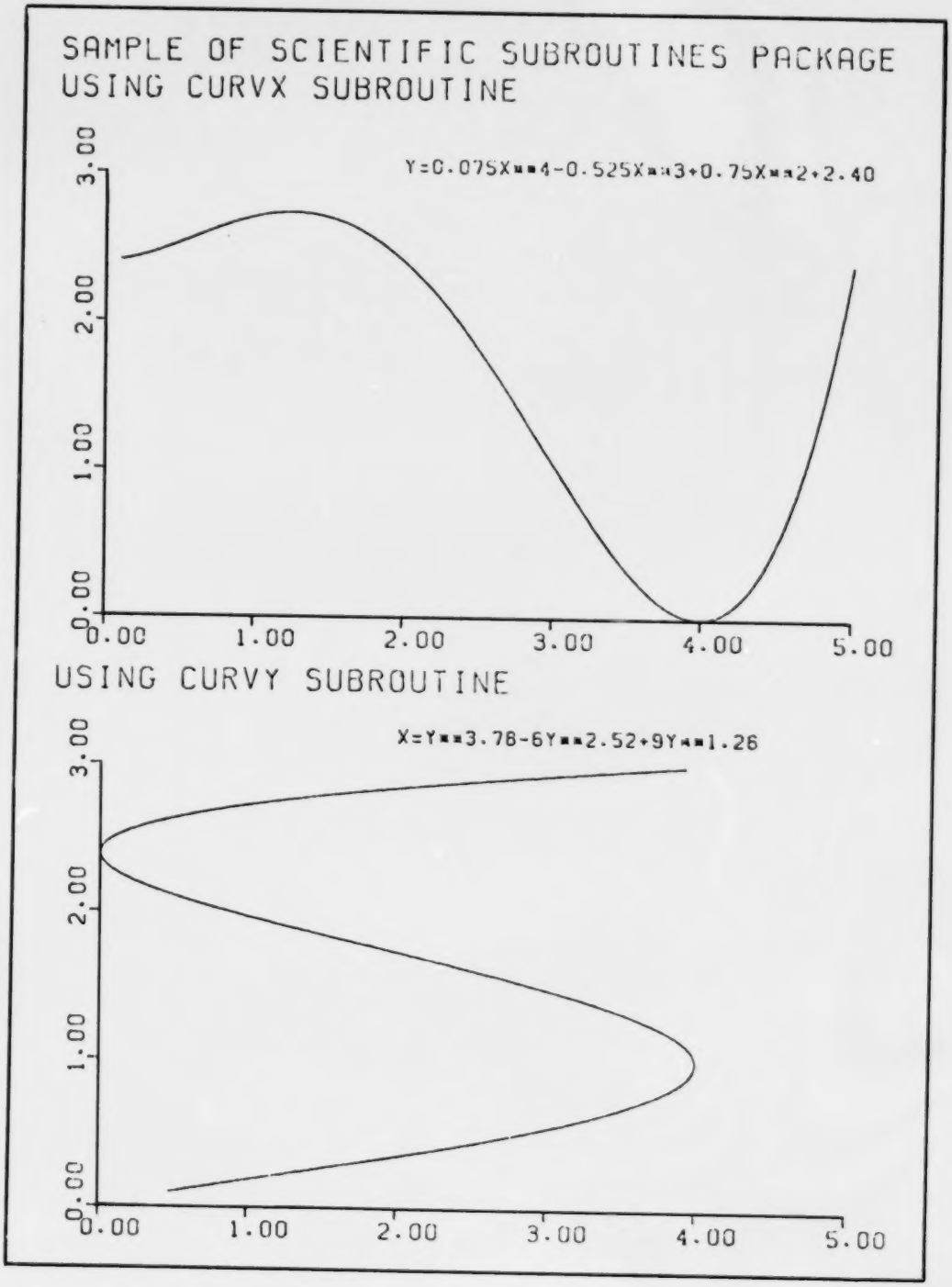


Figure 6
Samples of Scientific Subroutines Package

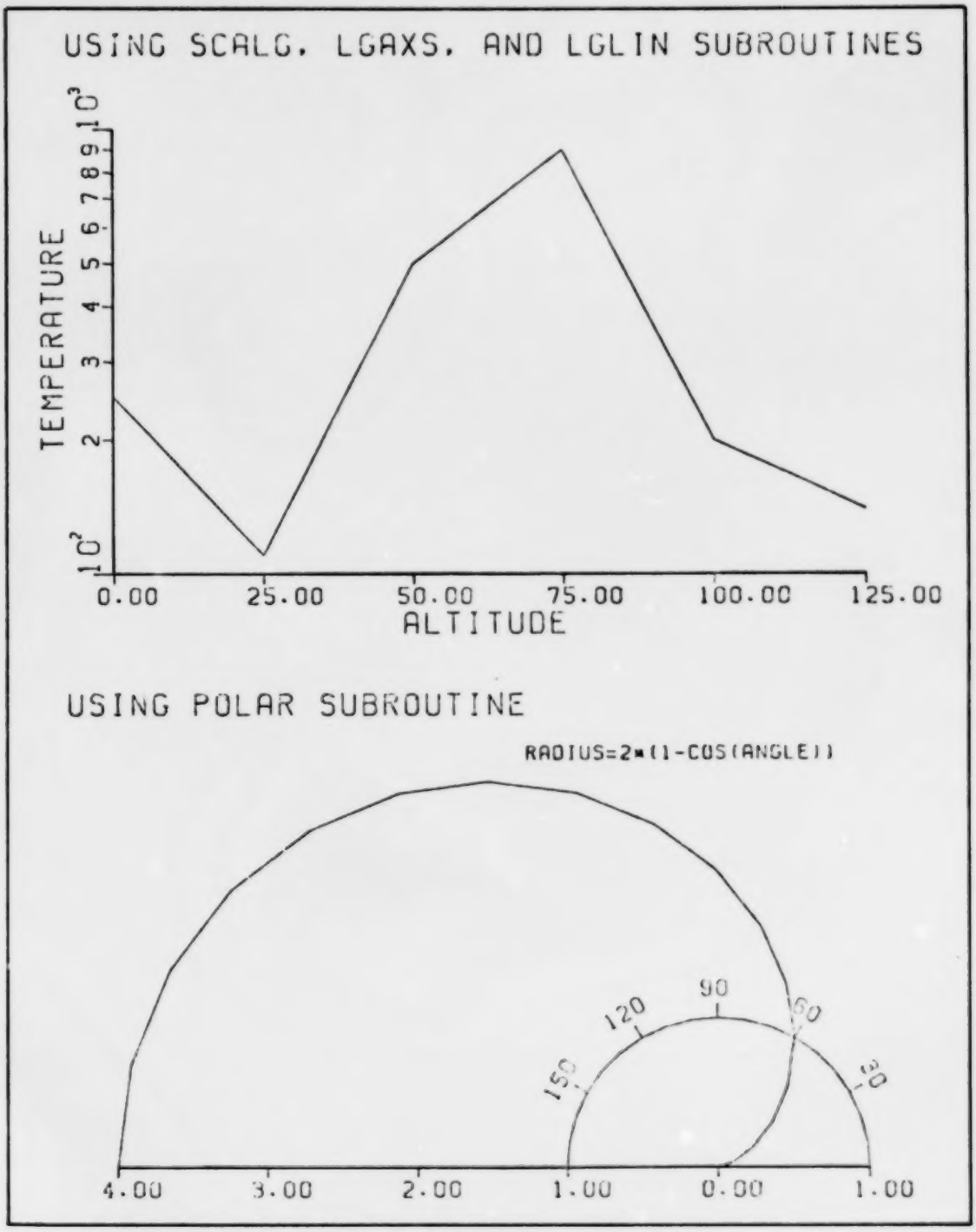


Figure 7
Samples of Scientific Subroutines Package

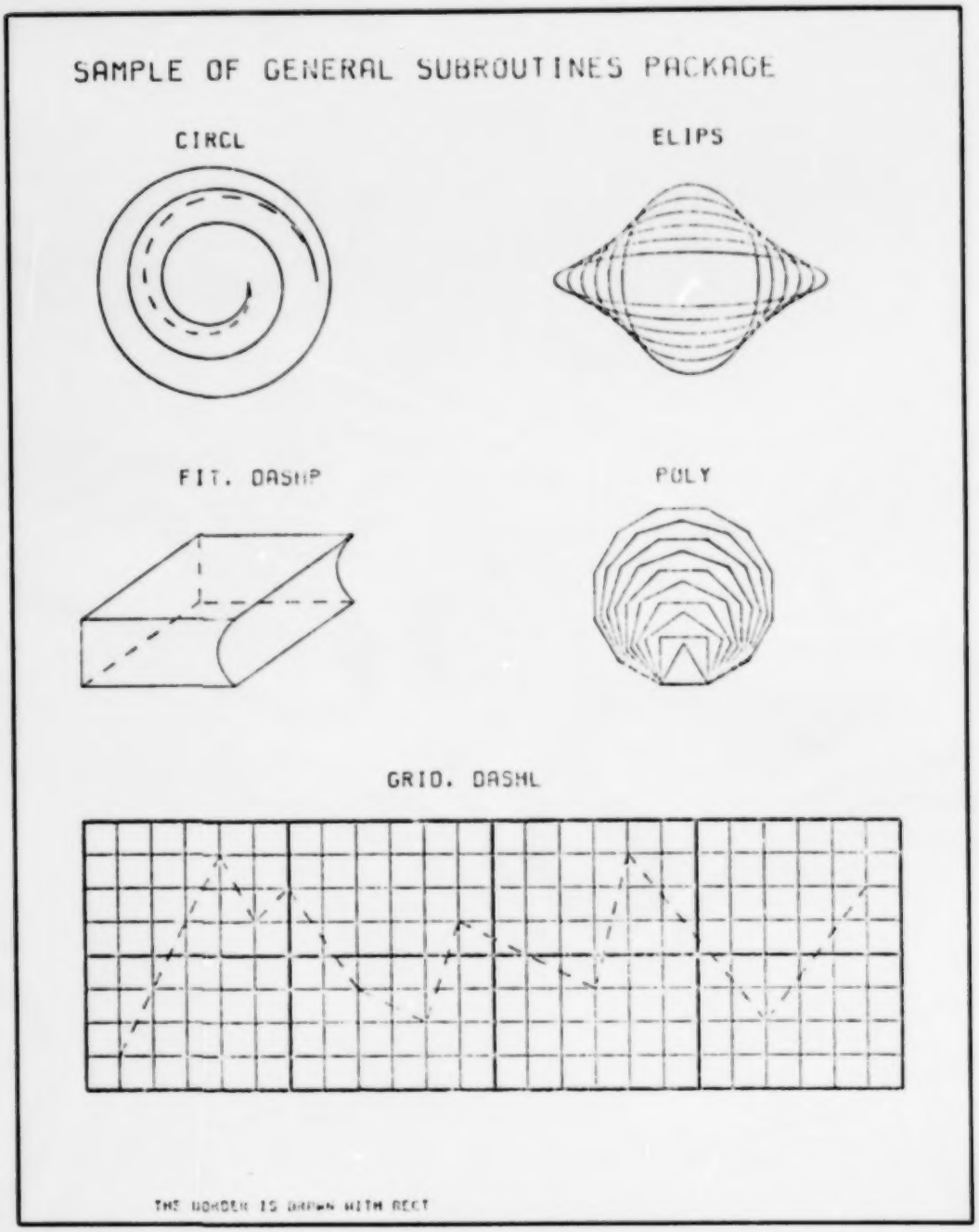


Figure 8
Sample of General Subroutines Package

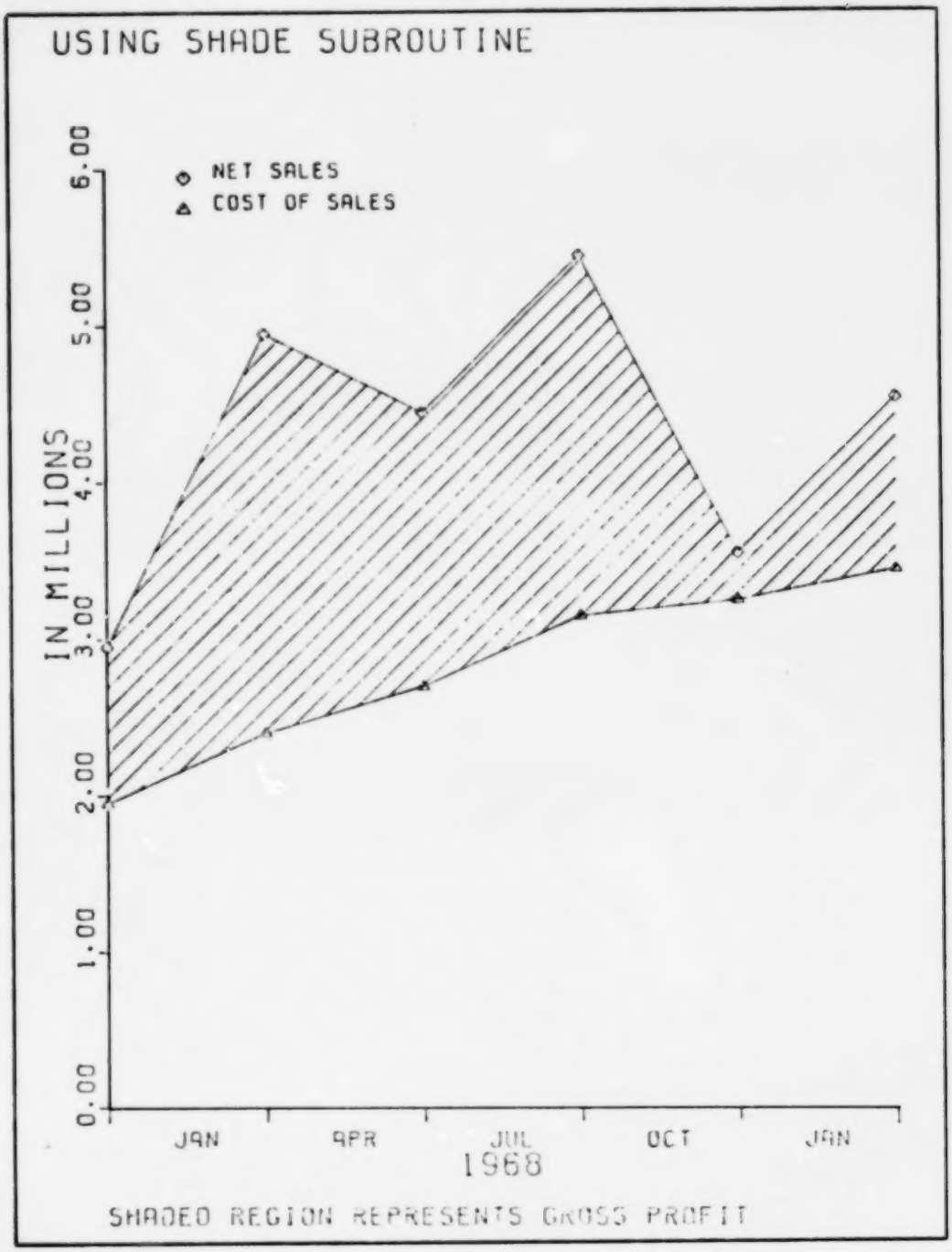


Figure 9
Sample of General Subroutine Package

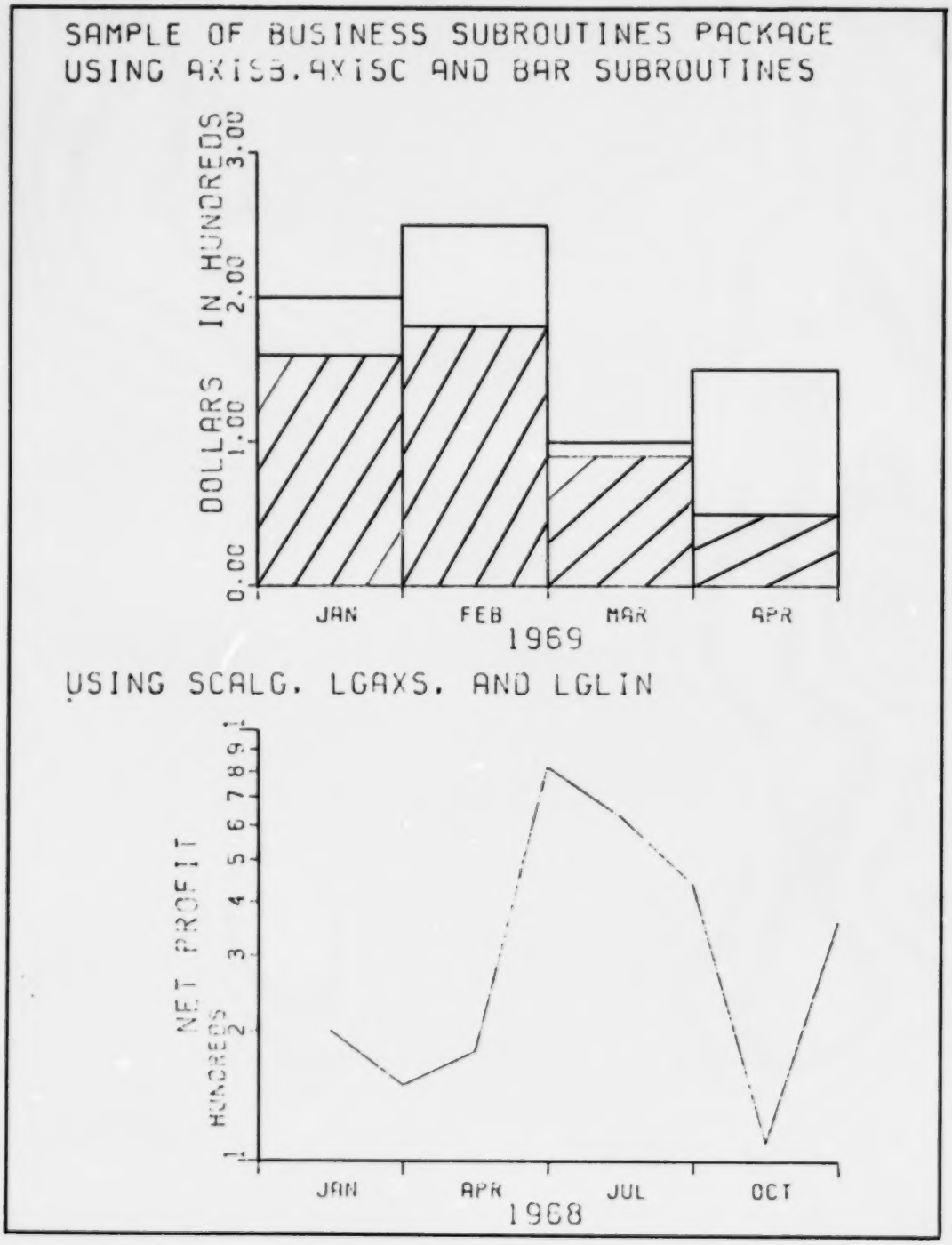


Figure 10
Samples of Business Subroutines Package

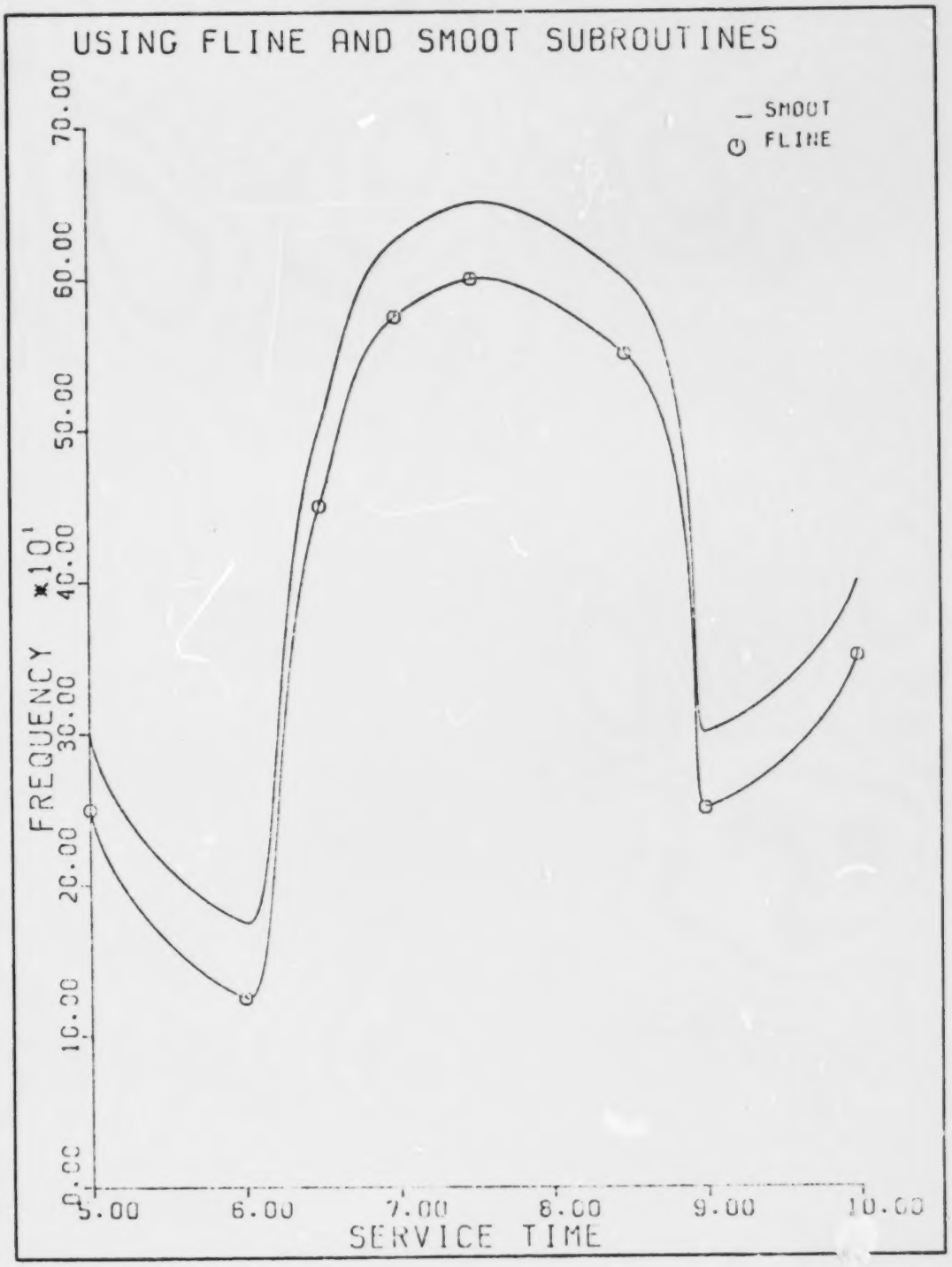


Figure 11
Sample of Business Subroutines Package

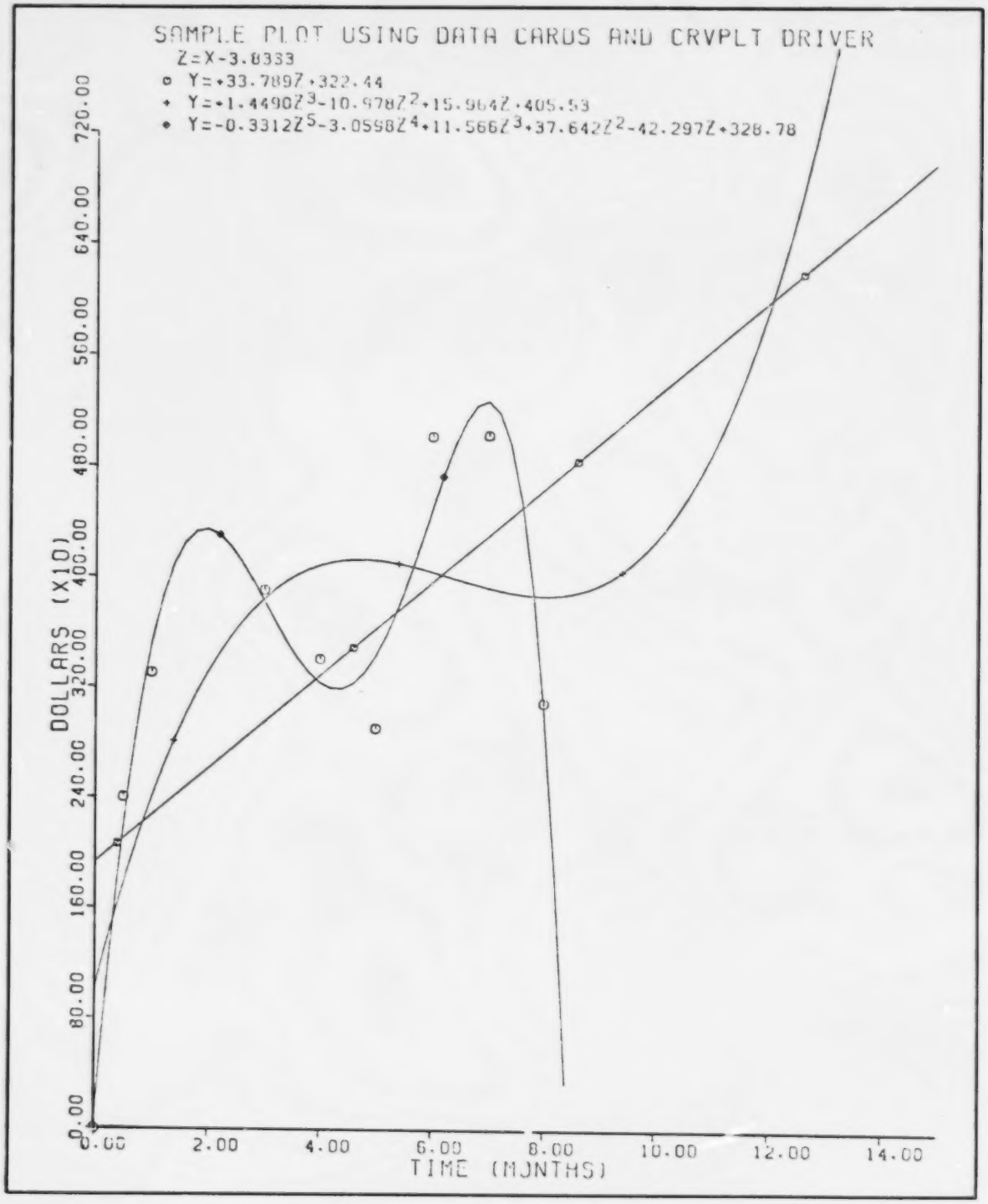


Figure 12
Sample Plot Using Data Cards and CRVPLT Driver

CALCOMP FUNCTIONAL
PLOTTING SUBROUTINES

AXIS
LINE
SCALE
SYMBOL
NUMBER
LBAXS
SHADE
STACK
AROHD
ARROW
CNTRL
DIMEN
LABEL
CURVX
CURVY
FIT4
FLINE
LGAXS

LGLIN
POLAR
REFLX
SCALG
SMOOT
CIRCL
DASHL
DASHP
ELIPS
FIT
GRID
POLY
RECT
SOLUT
AXISB
AXISC
BAR
CRVPT

Figure 13
CALCOMP Functional Plotting Subroutines

HARVARD GRAPHICS PACKAGES

By Curtis S. Lackey

PREFACE

This paper consists of excerpts of the systems manuals of the subject programs. Permission has been obtained to reprint these excerpts from the Laboratory for Computer Graphics and Spatial Analysis at Harvard University and from Professor Elliot E. Dudnik at the University of Illinois at Chicago Circle.

Interested parties should obtain a copy of LAB-LOG from the Laboratory. Address inquiries to:

Laboratory for Computer Graphics & Spatial Analysis
Graduate School of Design Harvard University
520 Gund Hall
48 Quincy Street
Cambridge, Massachusets 02138

LAB-LOG summarizes the materials available from the Laboratory and details ordering procedures.

INTRODUCTION

The Laboratory for Computer Graphics was established in the Spring of 1965 with a grant from the Ford Foundation to the Graduate School of Design of Harvard University. Under the direction of Howard T. Fisher, and with the advice and strong support of Dean Jose Luis Sert and Professor William Nash, the Laboratory has developed programs for high-speed electronic digital computer mapping and new techniques for graphic display that utilize the accuracy, thoroughness, speed, and low cost of computers. Originally based in the Department of City and Regional Planning, the Laboratory during the past two years has become a service organization to all departments in the Graduate School of Design. Recent projects have included faculty and students from each of the departments and programs, and currently every department shares in the Laboratory's activities through joint appointments.

Research in the Laboratory is essentially of two types. The first is a continued investigation into the uses of graphical analysis, and computer graphics in particular. Graphics is often a desirable mode of analysis, synthesis, and communication, and in the past, manually executed display has been adequate for many situations. Today's problems, however, are so complex that many elements of great diversity must be weighted in order to consider even the most rudimentary of them. Graphic display must be capable of a new level of sophistication and reliability in order to provide a sound basis for planning decisions of all kinds. Parenthetically, it is important to note that the Laboratory has not been concerned with the development of computers as machines. Rather, its aim has been to develop applications for existing, widely available equipment.

The Laboratory's past work in graphics was built largely on the computer mapping program developed at Northwestern University's Technological Institute by Howard T. Fisher. The technique, called the Synagraphic Mapping System, or SYMAP, is capable of composing spatially-distributed data of wide diversity into a map, a graph, or other visual display. The Laboratory initially built its work on the SYMAP program because of its great adaptability and because it uses the most widely available kinds of computer equipment. However, investigation has also been directed to other types of programs and to applications of cathode ray tubes and line plotters.

The second type of research undertaken by the Laboratory is pure research in the framework of general systems theory and spatial patterns. Much of the spatial analysis research constitutes work in what might be called "general spatial systems" theory, undertaken in relation to architecture, landscape architecture, city and regional planning, and urban design, with emphasis on the roles of computers in programming, design, simulation, and evaluation. In addition, a strong research effort has been organized in theoretical geography, supported at present by the Office of Naval Research. Here, there are three particular lines of endeavor: the theory of surfaces as related to spatial structure and spatial processes for geographical phenomena; the macrogeography of social and economic phenomena; and central place theory. Needless to say, these are not mutually exclusive categories, and already several major syntheses have been produced. The work is obviously undergoing a period of continued mathematization. Heavy use of the various geometries seems assured. Furthermore, the distinctions between various systematic branches of spatial study diminish at the theoretical level as common spatial properties are

recognized and as spatial solutions increasingly cut across traditional subject matter.

As a service organization to the Harvard Graduate School of Design, the Laboratory maintains data banks for several courses and assists in the establishment of computer programs and in the design of models for decision making. The Laboratory staff currently teaches nine courses within the School of Design and participates frequently in seminars. In addition, the Laboratory maintains out-of-house services to other universities, government agencies, and planning organizations. A prime policy of the Laboratory has been to disseminate its programs for the widest possible use; correspondence training in the use of SYMAP and OTOTROL has been and continues to be offered, and several intensive teaching sessions have been held at various universities and other organizations. The Laboratory has held two conferences, one for planners and geographers at Harvard in the Spring of 1967, the second, primarily for architects, at the Center for Continuing Education, University of Chicago, June 1968. A series of luncheon seminars dealing with many aspects of the Laboratory's interests were held at Harvard in the Springs of 1966, 1967, and 1968. Speakers were the Laboratory's staff and invited guests. Finally, a number of demonstration contracts have been undertaken, the largest current project being to show the advantages of computer mapping programs in air pollution studies.

The Laboratory was reorganized on July 1, 1968, as the Laboratory for Computer Graphics and Spatial Analysis and made a part of the Harvard Center for Environmental Design Studies within the Graduate School of Design. This move was a result both of the Laboratory's larger role as a computer consultant to the Graduate School of Design and of an expanded concept of its intellectual goals. William Warntz became Director of the Laboratory in July 1968.

In the immediate future, the Laboratory plans to continue its dual research and development emphases in computer graphics and spatial analysis. It hopes to further these objectives by strengthening the program of joint appointments within the Graduate School of Design and by providing more opportunities for graduate and undergraduate students to become involved in computer-assisted research. It further hopes to build upon the wider contacts already made with planners, designers, statisticians, systems analysts, economists, computer technicians, programmers, etc., both inside and outside the university.

THE SYMAP PROGRAM

SYMAP is a computer program for the production of maps and diagrams which graphically depict spatially disposed quantitative and qualitative information. It is suitable for a broad range of applications in a variety of disciplines and is provided with numerous options to meet widely varying requirements.

The term "synagraphic" is used to designate the general type of computer mapping to be described. It is derived from a combination of the words "synactic" and "graphic," the former an English word meaning acting together, cumulative in effect (from the Greek word "synagein," to bring together). The first syllable of SYMAP is pronounced as in symbol.

The program permits raw data of many kinds (physical, social, economic, etc.) to be related, manipulated, weighted, and aggregated in a variety of ways subject only to the user's needs or requirements. By assigning values to the coordinate locations of data points or data zones, one or more of three basic types of map may then be produced, as specified by the user.

The concept, overall design, and mathematical model for the SYMAP program were developed in the autumn of 1963 by Howard T. Fisher, working at Northwestern Technological Institute. Programming was carried out by Mrs. O. G. Benson of the Northwestern University Computing Center. Since that time many others have contributed ideas and cooperated in bringing the program to its present state. Much of the work has been accomplished at the Laboratory for Computer Graphics and Spatial Analysis established in 1965 at Harvard University. The present expanded program, SYMAP, Version 5, is the product of this facility with particular debts owed to Robert A. Russell and Donald S. Shepard.

GENERAL DESCRIPTION

The SYMAP program is written in FORTRAN IV. The source program consists of a main program and 49 subroutines and is approximately 5000 cards in length. At present, the program requires either a 200K partition if implemented without overlays or a 108K partition with overlays. The program has been successfully implemented on numerous computers ranging from the IBM 709 to the IBM 370. Comparable CDC, UNIVAC, XES, or HONEYWELL equipment have been or can be used with little or no modification.

TYPES OF MAPS

The three types of maps most commonly produced by the SYMAP program are:

1. *Contour* (based on the use of contour lines, each of which represents a uniform value throughout its length; B-DATA POINTS package). The contour (or isoline or isopleth) map consists of closed curves known as contour lines which connect all points having the same numeric value or height. Contour lines emerge from a datum plane at selected levels which are determined from the scale of the map and the range of the data. Between any two adjacent contour lines, a continuous variation or slope is assumed. Therefore, the use of contour lines should be restricted to the representation of spacially continuous information, such as most topography, rainfall and population density.

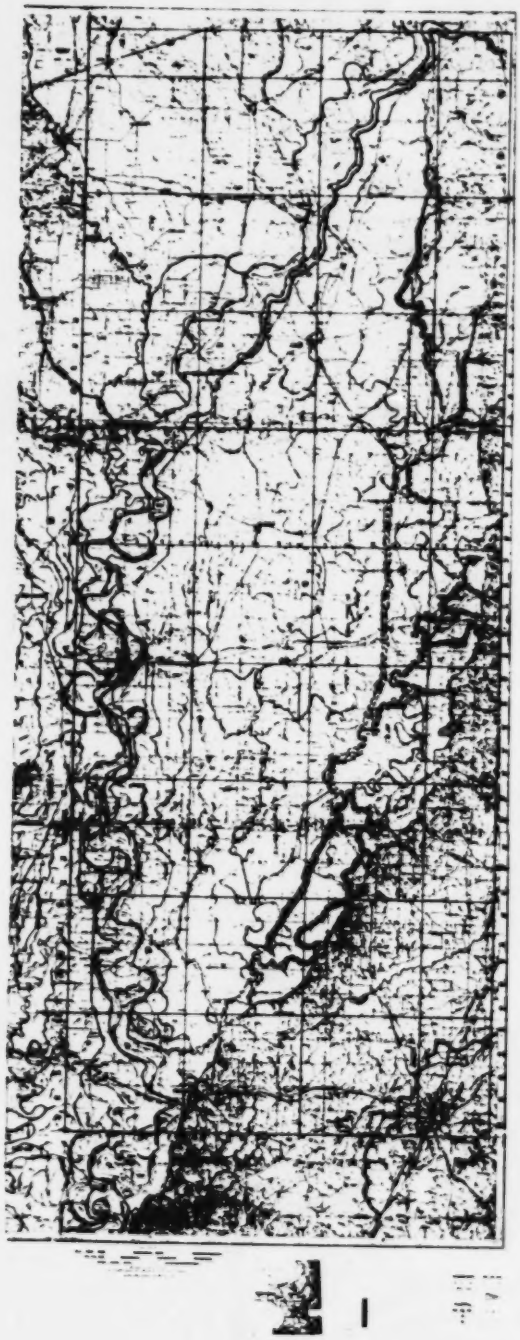
2. *Conformant* (based on conformance to the boundaries of a data zone; A-CONFORMOLINES package). The conformant (or choropleth) map is best suited for data, either qualitative or quantitative, whose areal limits are of significance, and whose representation as a continuous surface is inappropriate. Each data zone is enclosed by a boundary "conformant" to some predefined spatial unit. The entire spatial unit is given the same value, and symbolism is assigned according to its numeric class. Local variation of the data within the boundary will not be apparent, but on the average, will be correct.

3. *Proximal* (based on proximity to a data point; B-DATA POINTS package and electives 31, 36 and 37). The proximal map is very similar in appearance to the conformant map. However, the spatial units are defined by nearest neighbor methods from point information. Each character location on the output map is assigned the value of the data point nearest to it. Boundaries are assumed along the line where the values change and conformant mapping is applied.

Dissolved Oxygen

Coliforms

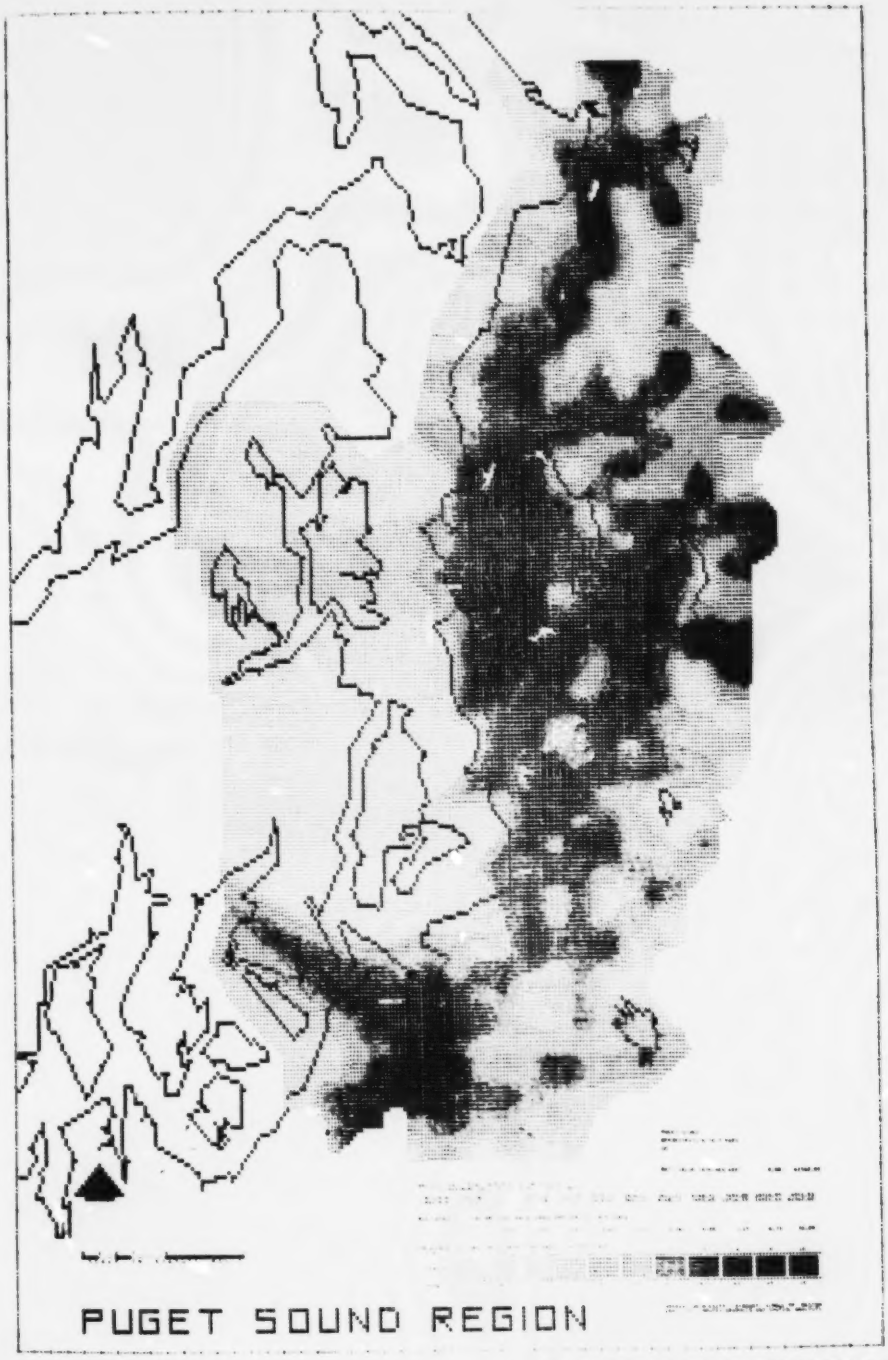
Figure 1
SYMAP



U.S.G.S. Map With
Grid Overlay

Average Rain During
Growing Season

Figure 2
SYMAP



PUGET SOUND REGION STUDY
Carbon Monoxide Motor Vehicle Emissions
Cities and Corridors Plan — System 35

Figure 3
SYMAP

THE SYMVU PROGRAM

SYMVU is a computer graphics program written for the purpose of generating three dimensional line-drawing displays of data. The SYMVU program is usable by persons with very little programming experience. Only three control cards are necessary for the generation of the graphic displays. Thirty-two electives or options are built into the program allowing for considerable flexibility in generating the displays of data. However, only seven of these options are absolutely necessary for the production of a single display.

SYMVU is most commonly used for quantitative geographic mapping purposes. Unlike contour mapping, SYMVU illustrates the absolute values of spatially continuous data. Contour maps (or isoline maps) are rounded intervals for mapping displays of quantitative information. In the case of SYMVU absolute values of a continuous surface are illustrated. The SYMVU program also has the capabilities of conformant and proximal mapping using data generated by the SYMAP program.

One of the distinct features of this program is that it accomplishes the task of deciding which parts of the object being viewed are seen and which are hidden from view.

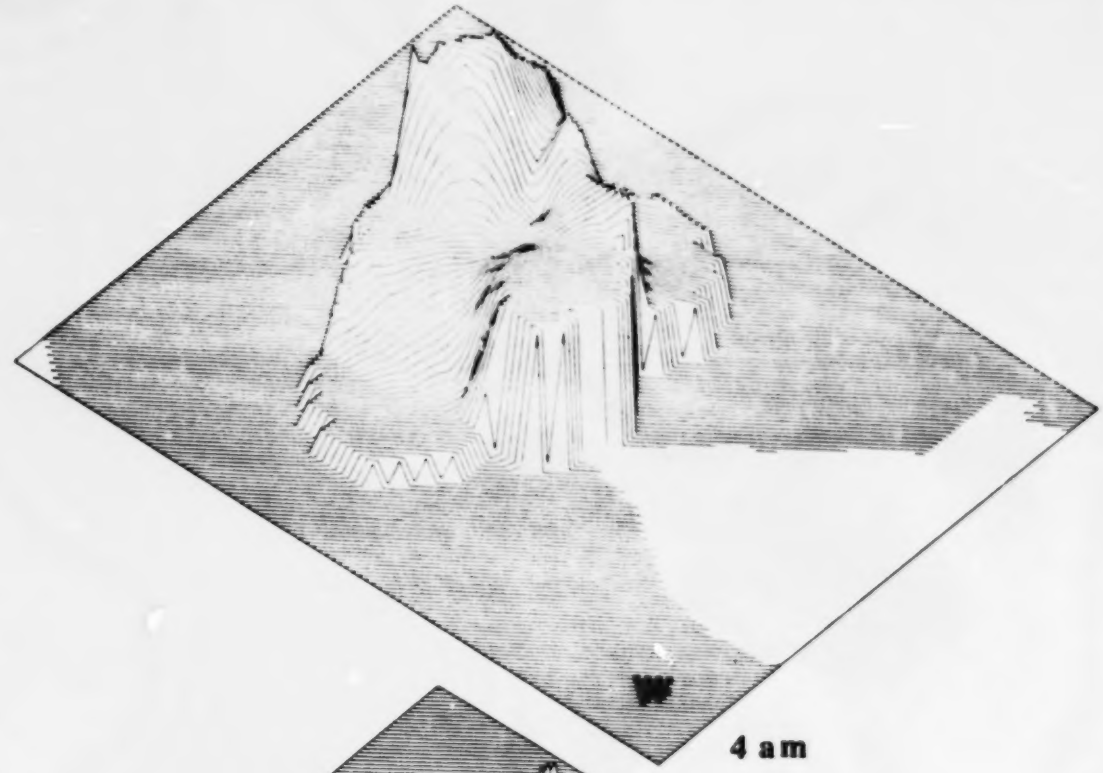
The program is written in FORTRAN-IV and is currently being operated on the IBM 360-65 using 220K memory. The output device that draws the illustrations is a CalComp plotter. Runs have been successfully completed on both an 11 inch and 30 inch plotter, although, because of a number of size constraints within the program a 30 inch plotter presently has little advantage.

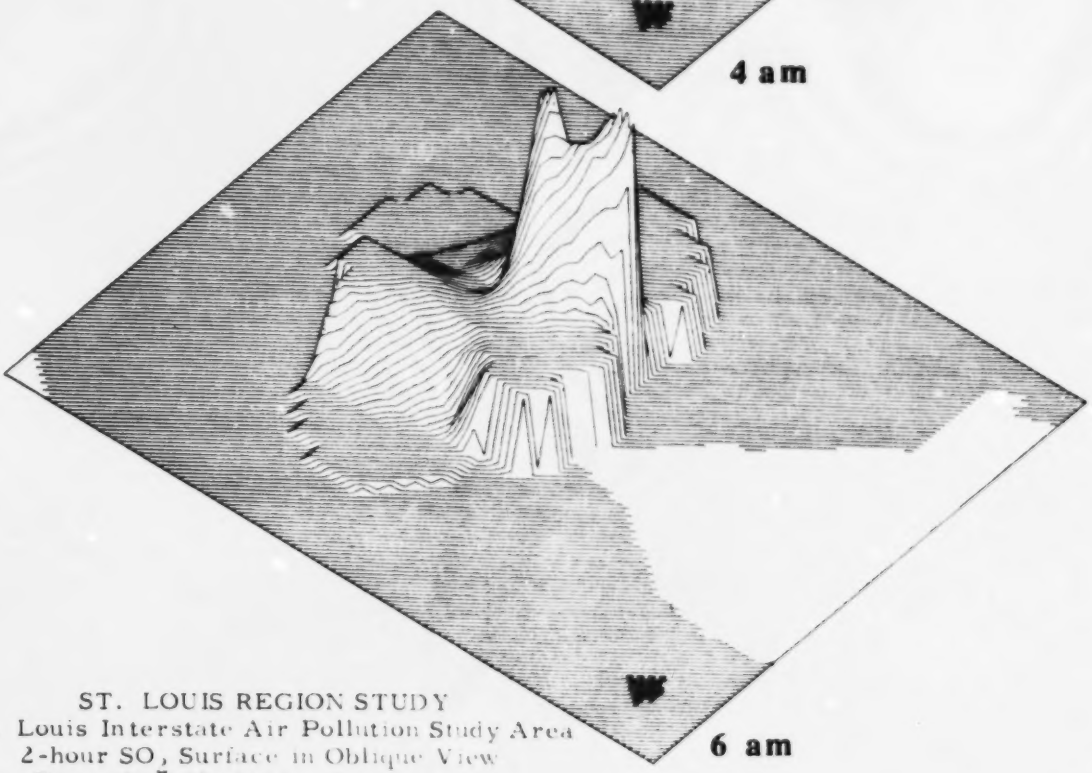
BASIC PRINCIPLES

The program utilizies basic data which is in the form of a grid matrix. This matrix information can be generated in either of two ways.

The first and most common alternative is to utilize data generated by the SYMAP program as a two-dimensional graphic display. SYMAP interpolates data from point or area locations so as to produce spatially continuous surfaces organized in the form of a grid matrix. This data can then be utilized by the SYMVU program for producing three-dimensional graphic displays.

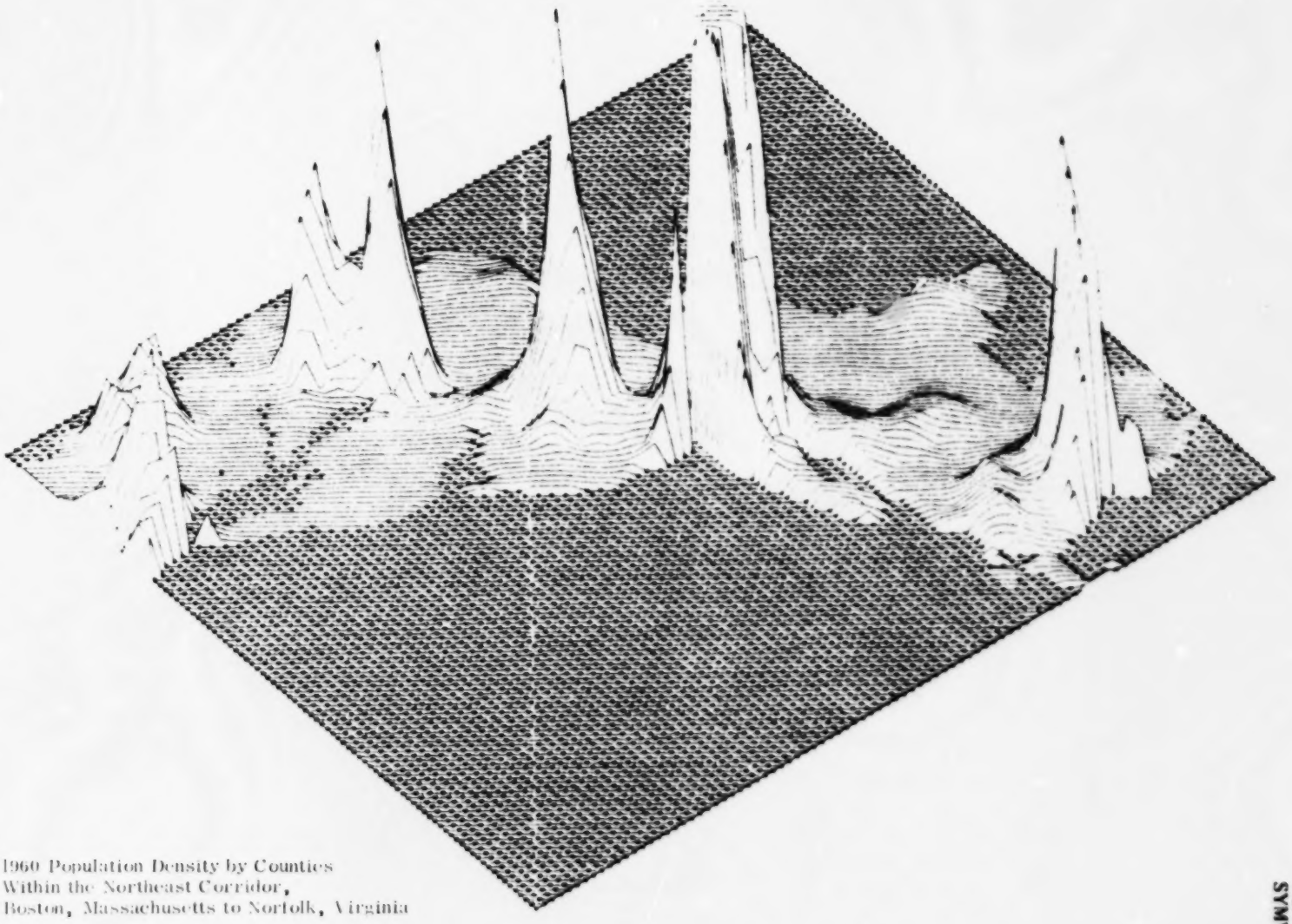
The second alternative involves providing data directly in a matrix form, either on cards or tape.





ST. LOUIS REGION STUDY
St. Louis Interstate Air Pollution Study Area
2-hour SO_2 Surface in Oblique View
December 20, 1964 — 4 and 6 a.m.

Figure 1
SYMVU



1960 Population Density by Counties
Within the Northeast Corridor,
Boston, Massachusetts to Norfolk, Virginia

Figure 2
SYMVU

THE GRID PROGRAM

GRID is a computer program which has been created to provide a highly efficient means for graphic display of information collected on the basis of a rectangular coordinate grid. The GRID program is designed for use by persons with very little programming experience. However, it is usually necessary for the user to specify his own data formats in subroutine FLEXIN, and this requires an elementary knowledge of FORTRAN IV.

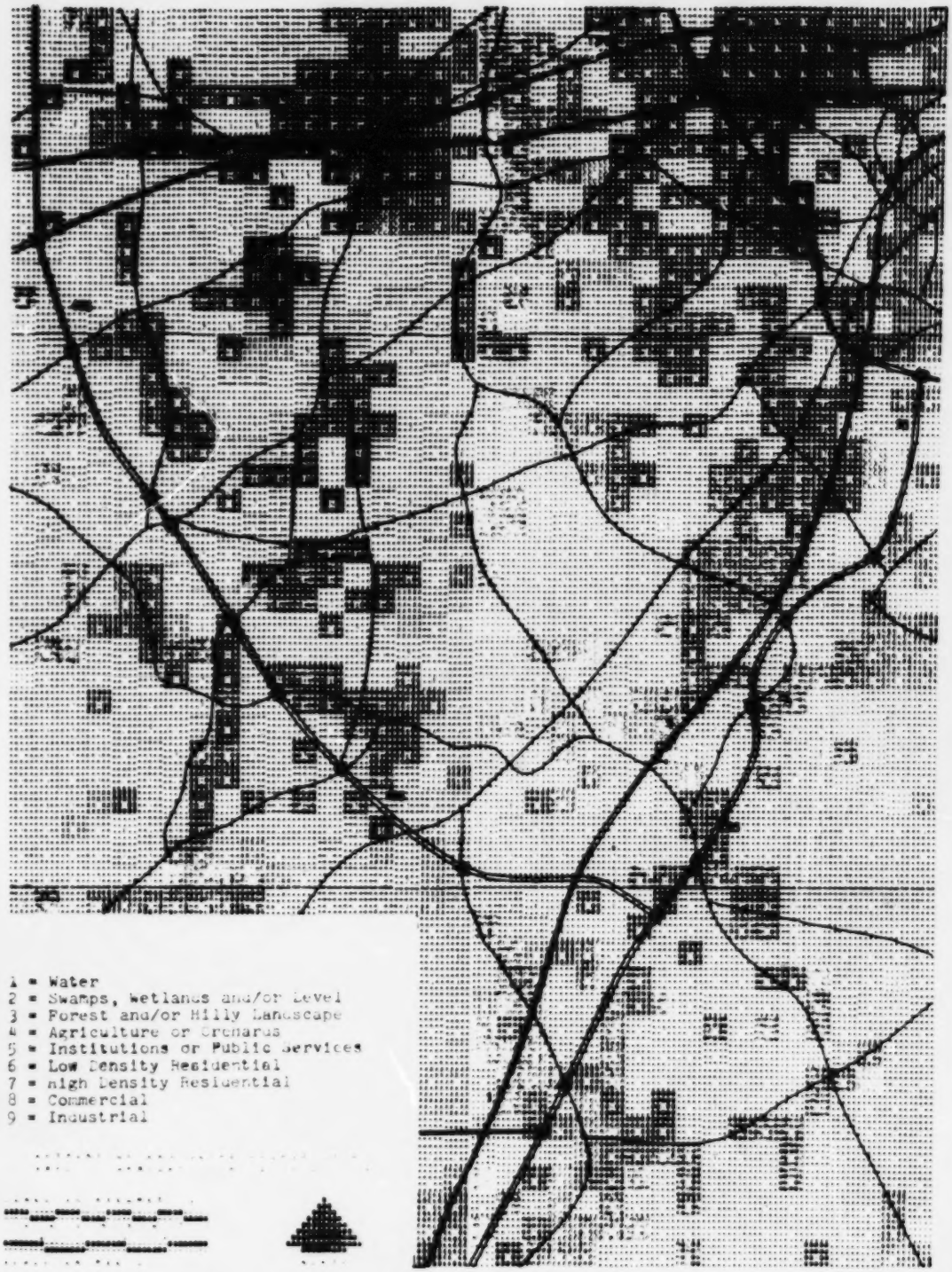
The program is written in FORTRAN IV and is currently being operated on an IBM 360/65 computer at Harvard using 150K bytes. With small programming changes it can be operated on the IBM 7094 with 32K memory. It is possible to operate this program on a smaller machine with a memory of at least 12K words.

BASIC PRINCIPLES

Each data value is assumed to be associated with a cell on a grid. It is essential that the values should be processed in the correct order, since the program accepts the data in the order in which it prints the map. By the standardized printing process, the program starts at the top of the map and processes the data horizontally row by row and from left to right in each row. Thus the numbers below represent the order in which thirty data values on a 6 x 5 grid will be processed and printed:

1	2	3	4	5	6
7	8	9	10	11	12
13	14	15	16	17	18
19	20	21	22	23	24
25	26	27	28	29	30

In the mapping process, the actual data values are generalized into groups, each group having a unique graphic symbol associated with it. The groups into which data are to be placed and the associated symbols may be specified by the user. Two types of symbolism are available: a grey scale of symbolism from light-dark which must be specified by the user, or a dot map in half-inch square cells. The coordinates of each grid cell location may also be printed. Most of the SYMAP electives for scale generation are available on GRID.



Predominant Visual Character, T = 0

Figure 1
GRID

THE CALFORM PROGRAM

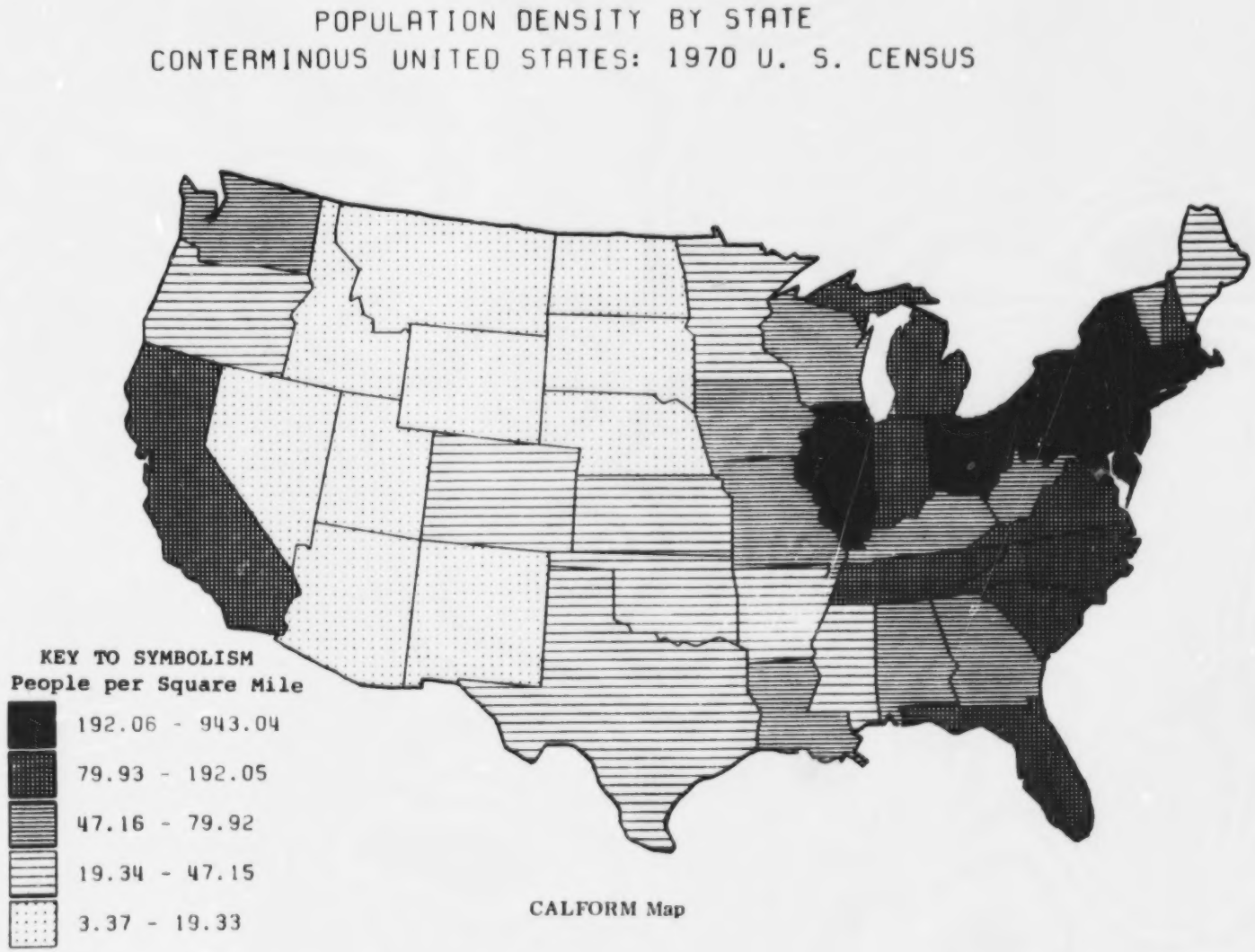
CALFORM is a computer program designed to produce conformant maps on a pen (or CRT) plotter. A conformant map is one in which symbolism (usually shading) represents the values of data attributed to data collection units (data zones) such as census tracts, municipalities, counties, and so forth.

Three steps are necessary to prepare a conformant map. These involve the definition of locations, values, and map options. Locations need be defined only once for a series of maps which portray various subjects for the same data zones.

The location of each data zone must be described as a series of straight line segments. Curved lines may be approximated by several straight line segments in order to preserve the degrees of detail desired. Straight line segments are defined in terms of the x-y coordinates at their end points. The resulting description of zone boundaries is called a computer readable base map or geographic base file. This data is punched onto cards and organized in the form of several functional packages.

The CALFORM program is written in FORTRAN IV and is currently operated on the Harvard/MIT IBM 370/165 using 128K bytes of core memory. The illustrations in this paper were produced on a CalComp Model 780, 11-inch incremental drum plotter.

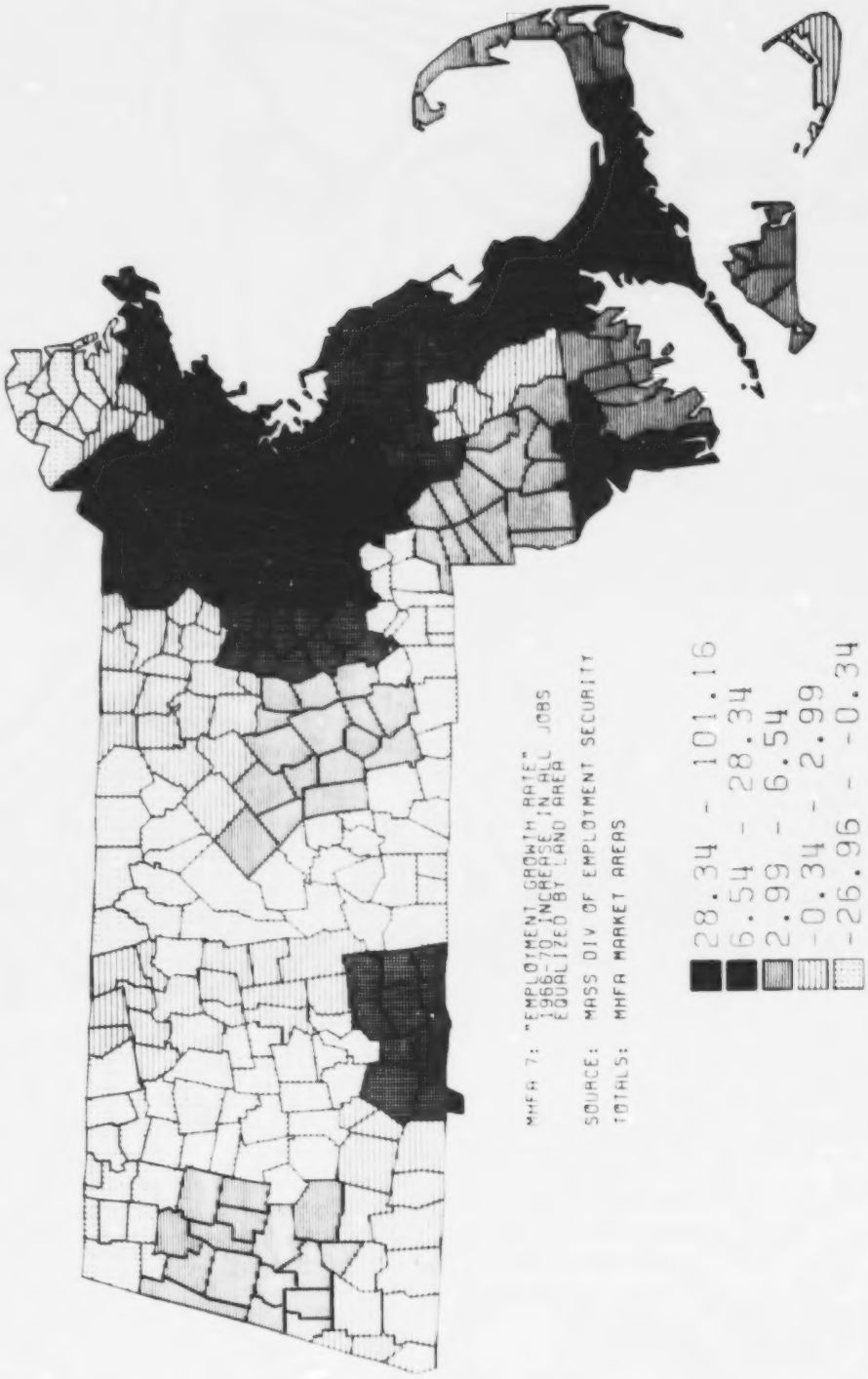
The program was written initially by Robert S. Cartwright, Jr., an applied mathematician and computer programmer for the Laboratory for Computer Graphics and Spatial Analysis, Harvard University. The original development of this program by Mr. Cartwright was sponsored by the Environmental Health Surveillance Center, University of Missouri in the summer of 1969. Since that time, the program and manual have been extensively revised at the Laboratory for Computer Graphics and Spatial Analysis. Individuals who have contributed to this work include Duncan Hughes, Kathleen Reine, David Sheehan, Geoffrey Dutton and Nicholas Christman.



CALFORM Map

Figure 1
CALFORM

Figure 2
CALFORM



THE POLYVRT PROGRAM

Recent years have witnessed an upsurge in the use of digital computers for geographical analysis and display, particularly in cartographic applications. Throughout industry, government, and academia, spatial data is being encoded, manipulated and displayed for an increasing variety of applications such as street address encoding, land use information retrieval, market analysis and cartographic display of social data. Due to the relative novelty and the diversity of uses of such applications, these computerized geographic base files (GBFs) currently exist in a variety of special formats, each suited to certain uses, and each requiring different computer software for their use. It may be many years, if indeed ever, before geographic information is encoded in a standardized fashion; some recording schemes are oriented toward fast selective retrieval of spatial data, other systems provide efficient storage for mapping purposes, while still others may sacrifice internal efficiency for ease of input and/or editing by inexperienced computer users. The lack of any common coding system for GBFs makes itself felt by restricting the exchange and use of existing data bases. This is particularly unfortunate in the case of massive base files, such as DIME files, which have been produced at great expense, and which may gather dust on users' shelves merely because the format is incompatible with the software required for the users' desired applications. And even when the file in question is not particularly cumbersome, it is still annoying at the least to have to recode the information merely because one wishes, for instance, to go from producing maps of the file on the printer to making plotted versions of those maps.

Without hoping for a universal GBF format to exorcise such bottlenecks, there is still no need for GBF users to recode their files by hand; this is the sort of job that computers have been designed to do, quickly, tirelessly and exhaustively. The missing link between different GBF formats exists now in the program POLYVRT (Polygon Converter), which can translate from any of several GBF types into any other of their formats, while at the same time allowing editing of the file, selective polygon retrieval, filtering out of unneeded line detail, coordinate transformations and scaling, and annotated plotted output of files as well as output in machine-readable form. POLYVRT does not translate between all different formats directly, but through a common data structure called a chain file which preserves all input information, as well as including information generated by POLYVRT for translation, such as left and right polygon designations which are lacking in most file formats. GBFs in chain file form may be written out in binary or formatted records and are acceptable input to the POLYVRT program itself. POLYVRT chain files may be created from or into the following GBF formats:

- DIME files (county, state, metropolitan)
- World Data Bank I chain files
- SYMAP Conformolines (output only)
- CALFORM Points and Polygons Packages.

More file types will be accommodated as the demand for them becomes known or the user may define his own file types to the program.

Although the chain file structure combines the advantages of many base file formats, it is not intended to replace them, except possibly as a medium for exchange of base files from one user's installation to another. Each input and output file type has its own strengths and special applicability, and through POLYVRT these may be most effectively exploited. Since the program edits, selects, checks for consistency and topological closure, generalizes overly detailed input, and allows visual inspection of files through plotted output, in addition to its conversion capacity, GBF users can be sure of obtaining with it the exact base file they need for virtually any application.

POLYVRT has a user-oriented command language organized into independent packages using fixed-field instruction formats. By the inclusion and exclusion of packages, all or only some of POLYVRT's capabilities can be invoked for any given execution of the program. Data may be read in with the input deck or from any sequential storage device. Some of the chain file features which POLYVRT produces are automatically stored off-line in temporary files, and this enables the program to manipulate files of nearly unlimited size in points. Using a carefully constructed list structure of pointers, core requirements are modest.

The POLYVRT program is the product of two years of design and development, during which time it has grown from a few routines to over 60 as new functions have been built into the system. It is not, however, the complexity of the programming, but the clarity of the data structures that brings an important new tool for computer cartographers and spatial analysts.

The data structure for Dual Independent Map Encoding (DIME) files arose out of the Census Use Study by the U.S. Bureau of the Census in preparation for the 1970 Census of Population and Housing and was developed under the direction of Caby Smith. The DIME

file (of U.S. County Boundaries) supplied with POLYVRT is a duplicate of the county DIME file distributed by the Bureau of the Census, reformatted, but not altered in structure.*

The central concept of POLYVRT's structure is the recognition of unbroken boundary units (chains) as the basic element of a geographic base file. World Data Bank I† the product of Warren Schmidt of the U.S. Central Intelligence Agency, provided a basis for the Laboratory's development. The need for easy access to a variety of data files formulated the requirements for a topological data structure. Most of the development of POLYVRT's file structure and the bulk of the code was the work of Nicholas Chrisman. James Little, author of the UPDATE package, also participated in the protracted testing and design process.

A particular debt is owed to David Douglas of Simon Fraser University, who with Dr. Thomas Peucker developed the detail filtering algorithm used in POLYVRT's F-GENERALIZE package, enabling superfluous or inconsequential detail in user's GBFs to be suppressed in files output from POLYVRT. The Douglas-Peucker algorithm was expanded by the Laboratory to allow a hierarchy of detail levels to be stored and retrieved by POLYVRT.

Constructive criticism and comments from many persons contributed to the process of programming POLYVRT, but specific acknowledgement to all might prove tedious to the reader. Special thanks is given to Allan Schmidt, Tom Peucker, Kathy Reine and Bob Fowler. Documentation was prepared by Geoffrey Dutton, carefully scrutinized and corrected by Nick Chrisman and Allan Schmidt.

REFERENCES

Dudnik, Elliott E., *SYMAP User's Reference Manual*, Chicago, University of Illinois.

Schmidt, Allan H., *LAB-LOG*, Cambridge, Harvard University, 1973.

Stinton, David and Steinitz, Carl, *GRID Manual*, Cambridge, Harvard University, 1971.

—, *CALFORM Manual*, Cambridge, Harvard University, 1973.

—, *POLYVRT User's Manual*, Cambridge, Harvard University, 1974.

—, *The Red Book*, Cambridge, Harvard University, 1970.

—, *SYMVU*, Cambridge, Harvard University, 1971.

*Copies of the county DIME files on tape, as well as copies of the metropolitan DIME files currently available, may be obtained from the User Service Staff, Bureau of the Census, Washington, D.C. 20233, for $70 per reel of tape. Certain metro DIME files require more than one reel.

†The World Data Bank File (version I, April 1972, in original format) is currently available from the National Technical Information Service, Springfield, Virginia 22151, as PB-2231178, on magnetic tape for $97.50 ($122.50 for foreign orders). Documentation for this file may be found in *Geographical Location Codes*, prepared by the General Services Administration of the Federal Government, Federal Supply Service number 7601-926-9078, available from the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402, for $2.75.

EPA'S REGIONAL AIR POLLUTION STUDY (RAPS)

By Robert B. Jurgens

DESCRIPTION OF THE REGIONAL AIR POLLUTION STUDY

The concept of a Regional Air Pollution Study (RAPS) was developed in the late 1960's. With direct White House support, RAPS became a reality when Congress agreed to allocate between $22 and $26 million to be spent over a 5-year period beginning in fiscal year 1973. The overall focal point and rationale for RAPS is the development of a series of Air Quality Simulation Models (AQSM) which can be used in least cost air pollution control strategies relative to EPA's current standards. These models could, for example, help in designing state pollution control implementation plans.

The criteria for choosing a region included having a city with significant air pollution, being isolated from influx of air pollution from surrounding areas and having a relatively simple topography. Of many candidate cities, St. Louis, Missouri, was chosen. The St. Louis region is a 20- by 14-kilometer area defined in Universal Transverse Mercator (UTM) coordinates.

The principal elements of RAPS are shown in Figure 1. As mentioned previously, the goal of RAPS is the development and evaluation of AQSM. Four functional modules comprise the sophisticated AQSM:

- An emissions model describing the location and temporal variations of emission rates of pollutant sources
- Atmospheric transport and diffusion
- Atmospherical chemical reactions of primary and secondary sources
- Removal processes such as rain, dry deposition or absorption by vegetation.

The RAPS emission inventory will encompass 12 separate source categories, the principal ones being illustrated. The individual methodologies are being designed for collecting emission inventories on an hourly basis. The data handling system for the RAPS emission inventories is scheduled to become operational in April 1975.

Paralleling the development of an emissions inventory is the establishment of a comprehensive aerometric data bank. The St. Louis Regional Air Monitoring System (RAMS) is a network of 25 fixed meteorological and air quality data acquisition sites. A Digital Equipment Corporation (DEC) PDP-8 is used for onsite process control and as a communications interface to the central facility. A PDP-11/40 at the central facility produces one 1200-foot reel of tape per day on which are recorded instrument calibration and status information as well as minute values and hourly averages of the aerometric data. The aerometric measurements that are made are shown in Table I.

Table I
RAMS Network Measurements

Meteorological		Air Quality	
Wind speed	WS	Sulfur dioxide	SO_2
Wind direction	WD	Total sulfur	TS
Turbulence	U	Hydrogen sulfide	H_2S
Temperature	T	Ozone	O_3
Temperature gradient	ΔT	Nitrous oxide	NO
Pressure	P	Nitrogen dioxide	NO_2
Dew point	DP	Methane	CH_4
Aerosol scatter (visibility)	b_{scat}	Total hydrocarbons	THC
Global solar radiation		Total suspended particulates	TSP
Direct solar radiation			
Terrestrial longwave radiation			

Augmenting the RAMS network are seasonal field expeditions where concentrated efforts are made in such areas as plume studies, wind field analysis, gas chromatography, elemental analysis, radiation and energy budget, aerosol programs, photochemistry and temperature-humidity effects on light scattering.

In addition to the ground based RAMS network, upper air measurements are routinely made. Meteorological balloon releases, pibals, and rawinsondes yield horizontal wind fields and vertical temperature profiles. During the field experiments, helicopter flights spiral down above certain field sites collecting a complex of aerometric data similar to that at the RAMS site below. Thus, a three-dimensional data bank for model development and evaluation is obtained. RAPS will be providing the most comprehensive emissions inventory ever amassed. It will be collecting the most comprehensive and best ambient air quality data ever obtained. The

meteorological conditions will be better described than ever before for an urban area, and all this data will be readily accessible.

This leads us directly to data management, the fourth element of RAPS. The 24-hours worth of RAMS network data which has been written onto magnetic tape at the St. Louis central facility is sent to the data management unit of the Meteorology Laboratory at Research Triangle Park.

A quick look survey program checks data consistency, reformats the data for archiving, creates mass storage data files of hourly data for inclusion into the RAMS S2K data base and produces a one-page summary printout. This summary information which can also be stored in a report file on the Univac 1110 computer includes station status information, diurnal meteorological conditions and key pollutant concentrations for selected rural and urban sites. This information will be displayed in the form of X-Y printer plots.

Access to minute values of RAMS data will be via sequential tape files where data for about 10 days can be stored on one 2400-foot reel at 1600 bytes per inch. Access to hourly data will be either via S2K immediate access mode for those having formal S2K training or through FORTRAN (or COBOL) programs. For browsing through the data base from a demand terminal, the immediate access mode would be used; whereas, FORTRAN programs, for instance, will interface the data base to the graphics package.

Data from the upper air network and from certain field experiments, which together with the RAMS data compose the RAPS data bank, will similarly be accessible via tape or S2K data base.

RAPS GRAPHIC SYSTEM

The maximum daily volume of hourly RAMS data is 12,000 words. The total RAPS volume per day is close to one million words. When dealing with a scientific data base of the size of RAPS, effective computer graphics become the primary mode of visual analysis and presentation of results.

The principal hardware/software support for graphics at Research Triangle Park is the UNIVAC 1110 facility which will support both off-line Calcomp plotting and interactive graphics terminals such as the Tektronix 4010. The Meteorology Laboratory has purchased two 4010's and associated hard copiers. With the aid of the Data Systems Division, the Laboratory has eliminated almost all bugs in the UNIVAC 1110/Tektronix interface. Transmission rates are now 300 baud but are expected to go to 1200 baud when some UNISCOPE 100 ports are freed. The Meteorology Laboratory owns a DEC PDP-11/40 disk operating minicomputer with a refresh graphics terminal. Data transmission rates for this cathode ray tube (CRT) are 9600 baud, thus permitting the possibility of movie generation at a later time.

Two task orders entitled "Computer Graphics Planning and Support" and "Plan for Interface between Graphics Display and RAPS Data Base" have been completed. The implementation phase is now in progress.

The structure of the interface between Data Base and Graphics Display is illustrated in Figure 2. The key to the system is the command program. The command language will be developed for both graphics and data requests. The same command language will be used on all RAPS data and graphics systems. Consequently, if the same set of commands are entered on punched cards or typed into a remote typewriter terminal, the results should be the same.

The description of the command language, the class of commands to be developed, and an example request illustrating the hierarchical command structure are given in Appendix A. These are from Dr. R. H. Allen's "RAPS Data Base and Graphics Interface Plan."

The RAPS Graphics Specifications are included in Appendix B and are taken from a Request for Proposal (RFP) of Graphics System.

CURRENT STATUS

The RAMS network is about 2 months from the 30-day acceptance period. The upper air network is scheduled to become operational November 1, 1974. The RAPS graphic system should be developed by the end of the year.

APPENDIX A - THE COMMAND PROGRAM

User and Command Program Interface

The command program is the FORTRAN main program that communicates with the user. All user requests are entered into the command program and all results

come back to the user through the main program. The main program has the following functions:

- Log user on
- Keep use statistics
- Prompt user
- Accept requests
- Give command error message
- Call data, graphics, utility, and library subroutines
- Return results
- Generate displays
- Keep status.

The standard interface between the user and the command program is the command language. The same command language is used for all input devices and all output devices. Of course some input devices cannot generate some input requests and some output devices cannot display some results. For example in the card input batch mode the user cannot take action on an unexpected response to a request. Also interactive graphics requests cannot take place when line printer simulated plots are being generated. The command language is considered in the next section.

The user inputs are entered on the following media:

- Punched cards
- Typewriter keyboard
- Position input on a graphics terminal
- Digitizing tablet input.

The results are returned to the user on the following media:

- Line printer
- Typewriter
- Graphics terminal
- Electrostatic printer/plotter
- Calcomp Plots.

Command Language

A common command language should be developed for both graphics and data requests. The same command language should be used on all RAPS data and graphics systems so that if the same set of commands are entered on punched cards or typed into a remote typewriter terminal, the results should be the same. The format and procedures for data or graphics requests should be the same.

The command language should have the following characteristics:

- Hierarchical
  - Full instructions
  - Prompting
  - Short form
- Forgiving
  - Checking
- Well-documented
  - Users
  - Programs
  - Examples
- Common language
  - All input devices
  - All output devices
- Graphics language subset of data language
- Any command can be given at any time
- Always store status so that program or command can be restarted (RSTR) in event of failure or interrupt.

The classes of commands should include:

1. Log
   a. On
   b. Off
   c. Terminal characteristics
2. Information

3. Control

   a. Access rights
   b. Mode

4. Status

   a. Default

5. Menu

6. Graphic

7. Data

8. Reporting

9. Movies and videotape.

The data and the graphics commands should have the following form:

xxxx - command description.

The four-character mnemonic should be chosen to be meaningful and unique. Typically the first four characters of the command description can be used. Other mnemonic formation rules include drop one letter of double letters, dropping silent letters, and dropping vowels.

The command format should be free form. Commands and parameters should be separated by spaces and/or commas. The order of a command should be:

1. Command
2. Input parameters
3. Output parameters.

On interactive terminals, display prompting characters should be displayed when waiting for input: REQ<. A command should be read in 80A1 format and scanned in a conversion routine to avoid conversion errors. Scan from the end of the line to find the last character in the line. Convert numbers by comparison with a list of 10 digits and special charaters. Break the buffer into parameters in the main program.

The full instruction mode for an example request might take the following form:

THE CLASSES OF COMMANDS ARE:

LOG - . . .

INFO - . . .

CONT - . . .

STAT - . . .

MENU - . . .

GRAF - MENU OF GRAPHICS COMMANDS

DATA - . . .

RPRT - . . .

MOVI - . . .

ENTER REQUEST>GRAF

WFLD - . . .

WSTA - . . .

WROS - . . .

CNVT - PLOT CONCENTRATION VERSUS TIME OF A POLLUTANT AT A STATION FOR A TIME RANGE

.
.
.

ENTER REQUEST>CNVT

THE POLLUTANTS ARE CO, CO_2, SO_2, . . .

ENTER REQUEST>SO_2

THE STATIONS ARE NUMBERED 1 to 25. SEE MAP.

ENTER REQUEST>2,

etc.

The prompting form might take the following form:

CLASSES - LOG, . . ., GRAF, . . .

REQ>GRAF

GRAF COMMANDS - WFLD, WSTA, WROS, CNVT, . .
.

REQ>CNVT

POLLUTANTS - CO, CO_2, SO_2, . . .

REQ>SO_2

STATION 1 TO 25

REQ>2

START TIME Y D H

REQ>Y4 D144 H14

END TIME

REQ>H17

The short form might take the following form: REQ>CNVT SO_2 2 Y4 D144 H14 H17.

The command plots the concentration versus time of SO_2 from station 2 starting at 2 p.m. on the 144th day of 1974 and ending at 5 p.m. on the same day.

APPENDIX B - RAPS GRAPHICS SPECIFICATIONS

As a supplement to these specifications, refer to the "EPA RAPS Graphics Plan," by Flow Research Inc., written as part of a RAPS task order. This document shall be used as a guide in developing the plan specified in the Scope of Work and as a reference in augmenting these specifications.

The contractor shall develop, test, and install a package which includes software at the following levels:

1. *Command Program.*—Contractor shall design and develop a command language and program compatible with the RAPS data management software and existing EPA facilities.

2. *Basic Routines.*—Contractor shall use existing Calcomp, Tektronix, and other software available within EPA where possible. As an example, a program to display wind roses on a Calcomp plotter has been developed by the Meteorology Laboratory. This should be incorporated into the package. All routines shall be FORTRAN callable and be cataloged in the command program.

3. *System Modifications.*—Contractor shall identify system requirements and work with EPA computer specialists in resolving conflicts or in implementing and testing modifications to the data base software.

Provision shall be made for displaying the graphics on a variety of existing and planned EPA computer peripherals. The EPA Univac 1110 computer system is the focal point of all RAPS data base activities. Supported graphics devices include a Calcomp 900/1136 Plotting System (off-line), a Tektronix 4010 series interactive demand terminal, and standard printers for simulated plotting. These devices, because of their speed, are suited primarily to producing a limited number of high-quality picture frames. They will be used to satisfy over 95 percent of the RAPS graphics requests. There are applications requiring an extremely large number of time sequenced frames, namely movie and videotape production. In these instances, it will be necessary to transfer portions of the data base and graphics software to a computer system which supports faster graphical output devices. A planned PDP-11/40 minicomputer facility with a high-speed interactive GT-40 CRT and electrostatic printer/plotter will be used for movie generation. Where possible, graphics routines shall be compatible with both computers and all peripheral devices. The command programs, however, shall be implemented on the Univac 1110 system only.

The preceding paragraphs enumerate general specifications regarding the type of software and facilities involved. Tables B-1 and B-2 contain detailed break-outs of the displays to be developed and their associated specifications:

The graphics system shall be readily applicable to a wide variety of data. Foremost is the application to the RAPS. After accessing the data base, the user shall be able to apply the desired graphics program in a straightforward manner. The contractor shall design the software so that the graphics is semiautomatically scaled to the data to be displayed.

During the course of this project, the contractor shall frequently consult with the EPA Project Officer and his designated representatives. These people will be readily available to the contractor so that system incompatibilities can be resolved in an efficient manner.

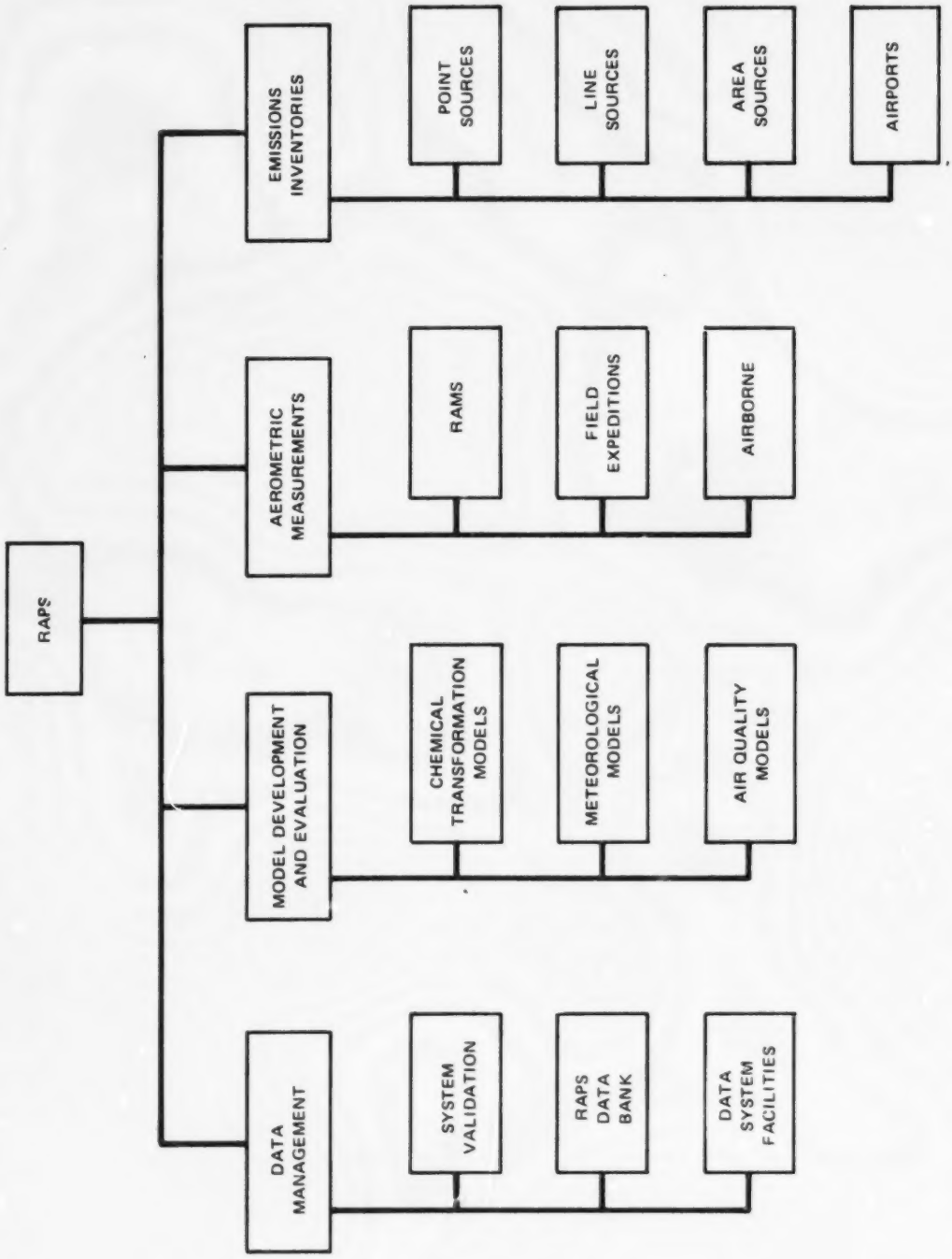
Table B-1
Displays to Be Developed

Display Types	Examples
Titles	
Text	
Plots and Graphs	
Graph of a function of one variable, f (x)	Concentration of pollution at a station over a time sequence
Family of curves of a function versus one variable at a list of of values of a second variable, f (x,y)	Concentration of pollution versus time at several stations
Multiple functions of one variable, f (x) and g (x)	Several pollutants versus time at one station
Contour plots, f (x,y)	Smoothed concentration contours as a function of geographic position
Three dimensional plots, f (x,y)	Terrain plots as a function of geographic position
Pictures	
Maps	Outline of St. Louis area
Field displays, amplitude	Wind data from RAPS stations
Grid field displays	Interpolated data at regular grid locations as a function of RAPS data
Engineering drawings	RAPS station layout
Charts	
Histogram and bar chart	Excess ozone concentration by stations
Pie charts	Sulfur emission percentage by source type
Flow charts	RAPS data management information flow
Wind roses	Wind direction and magnitude summaries for a RAPS station

Table B-2
Display Specifications

Characteristics	Graphics Commands
Distinguishable lines between variables. Weight, continuity, and color shall be used wherever supported by the output media.	STORE
Plotted symbols. Any size and orientation, on the character set which shall include A - Z, 0 - 9, and special characters (i.e., the Calcomp set).	PLOT
Annotation for graphs and plots.	SCALE: shall enable user to enlarge or compact display by a specified factor.
Shading	ZOOM: shall enable user to "blow up" designated portion of a display by a specified factor
Overlays	WINDOW: shall enable user to generate time sequenced displays.
Selectable graph scales shall include linear, semi-log, low-log, and combinations thereof.	MERGE: shall initiate overlap of displays.
Input positional data (where permitted by the device) to create, edit, and interact with displays.	EDIT

Figure 1
Principal Elements of Regional Air Pollution Study (RAPS)



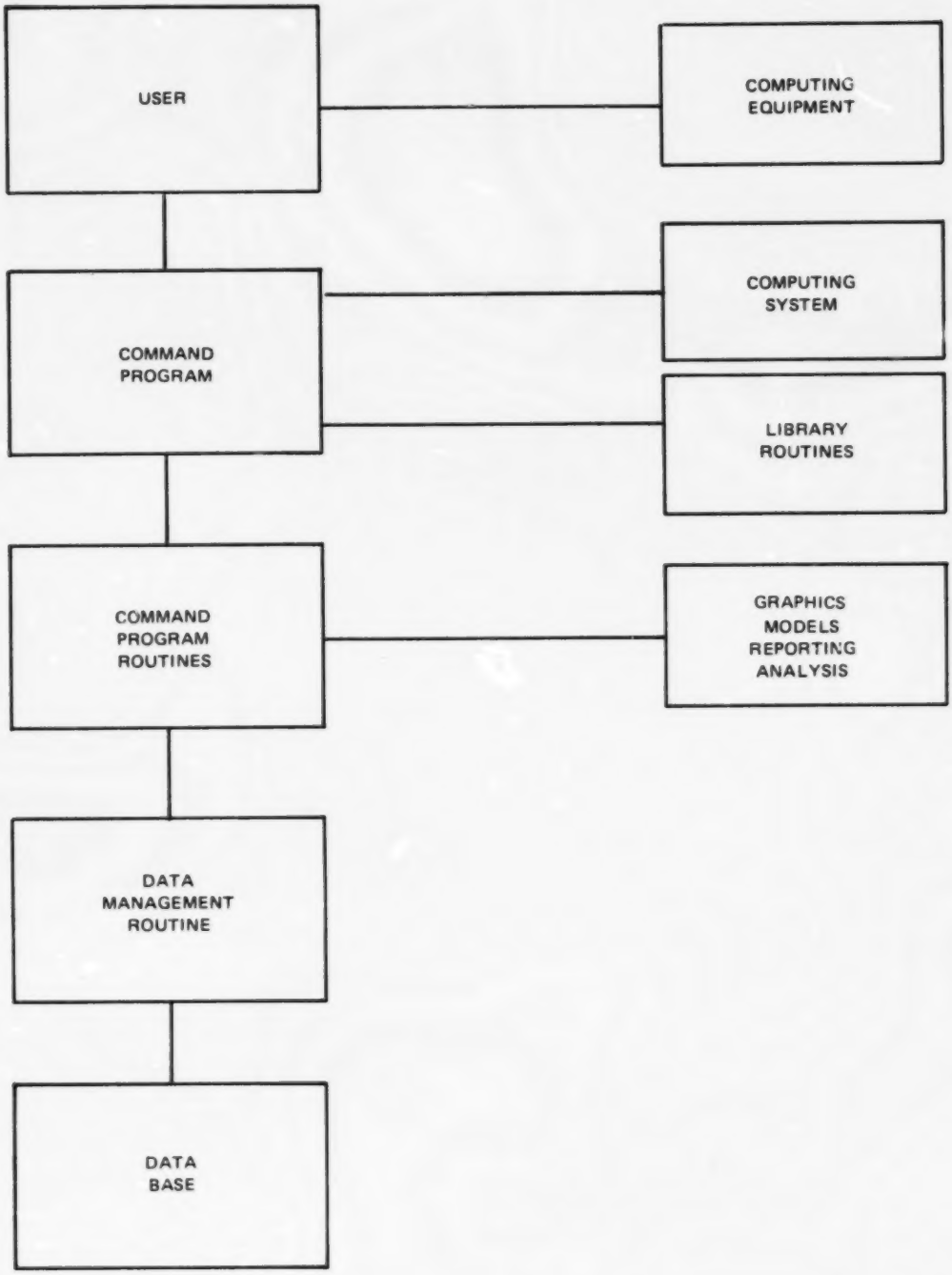


Figure 2
RAPS Data Base and Graphics System Interfaces

CAPABILITIES OF TEKTRONIX SOFTWARE

By David M. Cline

INTRODUCTION

Graphical display computer terminals have been in use for a number of years. Until recently, the display units were used essentially in a refresh mode and were, by necessity, connected directly to the computer that utilized them. With the advent of a low-cost graphic display storage unit, the use of graphics in a time sharing mode became very cost effective. The number of graphical display computer terminals that the Environmental Protection Agency (EPA) has acquired over the last two years indicates a commitment from the user community that graphical support is required.

As early as 1972, EPA personnel were utilizing Tektronix 4010 graphical display computer terminals at the Boeing Computer Services (BCS) IBM 360/67 computer facility at Wichita, Kansas, and at the National Institutes of Health (NIH) PDP-10 computer facility at Bethesda, Maryland.* Subsequently, the EPA contract with BCS was terminated in the spring of 1973, and it was not until late July 1973 that graphical support was available at the Optimum Systems Corporation (OSI) facility. The OSI graphic facilities were chosen because they are available to all EPA computer uses. The PDP-8/E graphic capabilities are discussed as they are available to a number of EPA installations.

GRAPHICS AT OSI

Graphical support at OSI is available on the IBM/360 model 155 and may be accessed via IBM's Time Sharing Option (TSO). The OSI systems group made modifications to the Telecommunications Access Method (TCAM) that were necessary for graphical support. Southeast Environmental Research Laboratory (SERL) personnel installed the Terminal Control System, the Advanced Graphing II, and the Calcomp Preview packages. The subroutine, TINPUT, which allows graphical input was in error as were several of the Advanced Graphing II subroutines; however, all were corrected before being made available as public files.

TERMINAL CONTROL SYSTEM

The Terminal Control System (TCS) package is the most fundamental of the packages and is used by the other graphics packages. It is written completely in FORTRAN IV in the form of 59 subroutines and is described in detail in the Tektronix document, 062-1474-00, entitled, "Terminal Control System - 4010, User's Manual." The subroutines are saved in a cataloged dataset, CNA 324.DMC.TCS.

It is assumed that the user has a knowledge of TSO to the extent that a FORTRAN dataset may be created and compiled. During the LINK or LOAD step, the library of subroutines may be accessed by the following type of command:

LINK your-dataset-name LIB('CNA324.DMC.TCS') FORTLIB. To utilize the subroutines, the FORTRAN files FT05F001 and FT06F001 must be allocated to your terminal. This is accomplished by issuing the following commands to TSO:

ALLOCATE DA(*) F(FT05001)

ALLOCATE DA(*) F(FT06001) or by executing the following TSO CLIST:

EXEC 'CNF324.DMC.FT5.CLIST' LIST (See Appendix.)

A TSO CLIST dataset is available to perform the process of compiling, linking, and executing of a FORTRAN dataset containing calls to the TCS package. It may be copied to your account by issuing the following TSO command:

COPY 'CNF324.DMC.TCS.CLIST' TCS.CLIST' (See Appendix.)

The CLIST is used as follows:

EXEC TCS 'dataset name' LIST

where the dataset name is the name of the FORTRAN dataset to be compiled without the FORT extension. If the dataset to be compiled and executed is DRAW.FORT, issue the following command:

EX TCS 'DRAW' LIST (See Appendix.)

*Mention of commercial products does not necessarily constitute endorsement by EPA.

ADVANCED GRAPHING II

The Advanced Graphing II (AGII) package is a sophisticated collection of FORTRAN subroutines that permits the graphing and labeling of data with little knowledge of the TCS package. The subroutines and examples for using them are described in the Tektronix document, 062-1530-00, entitled "Advanced Graphing II, User's Manual." The subroutines are saved in a cataloged dataset CNA324.DMC.AGII.

To use the AGII package, one must allocate the FORTRAN I/O files FT05F001 and FT06F001 as described above. A TSO CLIST is available to support the sequence of commands required to compile, link, and execute a FORTRAN dataset containing subroutine calls to the AGII library. It may be copied as follows:

COPY 'CNF324.DMC.AGII.CLIST' AGII.CLIST (See Appendix.)

The CLIST may be invoked as described above:

EXEC AGII 'dataset name' LIST (See Appendix.)

CALCOMP PREVIEW PACKAGE

The Calcomp Preview Package consists of a series of FORTRAN subroutines that are used in lieu of the standard Calcomp plotter subroutines to give the user the ability to preview the output of a Calcomp plotter on the Cathode Ray Tube (CRT). Thus, the user may view his proposed graph immediately with the interactive Tektronix software. This capability should prove to be of tremendous benefit to Calcomp plotter users in terms of decreased program development time. The subroutines, with examples for using them, are described in the Tektronix document, 062-1526-00, entitled "Preview Routines for Calcomp Plotters, User's Manual."

As with the other Tektronix packages, the FORTRAN I/O files must be allocated before using the Calcomp Preview routines. A TSO CLIST is available which contains the sequence of commands required to compile, link, and execute a FORTRAN program containing subroutine calls to the Calcomp replacement routines.

It may be copied by issuing the following TSO command:

COPY 'CNF324.DMC.CAL.CLIST' CAL.CLIST (See Appendix.)

Line 40 of this dataset must be modified to reflect the name of the dataset containing the Calcomp subroutines normally used for plotting. Line 40 of the dataset CAL.CLIST is listed below. The name of the dataset that must be replaced by the name of the user's dataset is underscored:

00040 LINK &NAME. ('CNF324.DMC.-CALPREV.LOAD','<u>CNF324.DMC.CALMIC.-LOAD</u>', 'CNA324.DMC.TCS') FORTLIB

Then, to execute a FORTRAN program that the user may wish to preview, issue the following command:

EXEC CAL 'dataset name' LIST (See Appendix.)

GRAPHICS ON A PDP-8/E

The PDP-8/E minicomputer system that supports the Aquatic Ecosystem Simulator (AEcoS) at the Southeast Environmental Research Laboratory (SERL) is equipped with a Tektronix 4010 computer display terminal, the manufacturer's OS/8 software system, and a FORTRAN IV software system. A Calcomp plotter for the computer system has been ordered to complement the graphics effort at SERL. Compiling the TCS subroutines and writing two machine dependent subroutines, TINPUT and TOUTPUT, were all that was necessary to implement the TCS package on the PDP-8/E computer. Upon arrival of the Calcomp plotter and associated software, it will be possible to preview Calcomp plots on the 4010 by simply linking to the Calcomp software replacement routines in much the same manner as on the OSI facility.

The TCS and Calcomp Preview routines that execute on the PDP-8/E can be easily exported to other PDP-8/E's; but the OS/8 and FORTRAN software cannot, since they were purchased on a licensing agreement on a per machine basis. As the work load at SERL permits, an attempt will be made to install the AGII routines on the PDP-8/E.

APPENDIX - LISTING OF TSO GRAPHICS CLISTS

Listing of FT5. CLIST

```
00010 ALLOCATE DA(*) F(FT05F001)
00020 ALLOCATE DA(*) F(FT06F001)
00030 END
```

Invoking FT5. CLIST

```
EXEC FT5 LIST
ALLOCATE DA(*) F(FT05F001)
ALLOCATE DA(*) F(FT06F001)
END
```

Listing of TCS. CLIST

```
00010 PROC 1 NAME
00020 FORT &NAME. NOPRINT
00030 WHEN SYSRC(GE 8) END
00040 SCRATCH &NAME . . LOAD
00050  LINK  &NAME.  LIB('CNA324.DMC.TCS')
FORTLIB
00060 WHEN SYSRC(GE 8) END
00070 SCRATCH &NAME . . OBJ
00080 CALL &NAME.(TEMPNAME)
00090 END
```

Invoking TCS. CLIST

```
EXEC TCS 'DRAW' LIST
FORT DRAW NOPRINT
G1 COMPILER ENTERED
SOURCE ANALYZED
PROGRAM NAME = MAIN
* NO DIAGNOSTICS GENERATED
WHEN SYSRC(GE 8) END
SCRATCH DRAW.LOAD
LINK DRAW LIB('CNA324.DMC.TCS') FORTLIB

WHEN SYSRC(GE 8) END
SCRATCH DRAW.OBJ
CALL DRAW(TEMPNAME)
```

Listing of AGII. CLIST

```
00010 PROC 1 NAME
00020 FORT &NAME. NOPRINT
00030 WHEN SYSRC(GE 8) END
00040 SCRATCH &NAME . . LOAD
00050  LINK  &NAME.  LIB('CNA324.DMC.TCS',
'CNA324.DMC.AGII') FORTLIB
00060 WHEN SYSRC(GE 8) END
00070 SCRATCH &NAME . . OBJ
00080 CALL &NAME.(TEMPNAME)
00090 END
```

Invoking AGII. CLIST

```
EXEC AGII 'WILL' LIST
FORT WILL NOPRINT
G1 COMPILER ENTERED
SOURCE ANALYZED
PROGRAM NAME = MAIN
     NO DIAGNOSTICS GENERATED
WHEN SYSRC(GE 8) END
SCRATCH WILL.LOAD
DATA SET WILL.LOAD NOT IN CATALOG
LINK  WILL  LIB('CNA324.DMC.TCS',  'CNA324.-
DMC.AGII') FORTLIB

WHEN SYSRC(GE 8) END
SCRATCH WILL.OBJ
CALL WILL(TEMPNAME)
END
```

Listing of CAL. CLIST

```
00010 PROC 1 NAME
00020 FORT &NAME. NOPRINT
00030 WHEN SYSRC(GE 8) END
00040 SCRATCH &NAME . . LOAD
00050  LINK  &NAME.  LIB('CNF324.DMC.-
CALPREV.LOAD',  'CNF324.DMC.CALMIC.-
LOAD', 'CNA324.DMC.TCS') FORTLIB
00060 WHEN SYSRC(GE 8) END
00070 SCRATCH &NAME . . OBJ
00080 CALL &NAME . . LOAD(TEMPNAME)
00090 END
```

Invoking CAL. CLIST

```
EXEC CAL 'DEMO' LIST
FORT DEMO NOPRINT
G1 COMPILER ENTERED
SOURCE ANALYZED
PROGRAM  NAME  =  MAIN 001  DIAGNOSTICS
GENERATED, HIGHEST  SEVERITY CODE IS 4
WHEN SYSRC(GE 8) END
SCRATCH DEMO.LOAD
LINK   DEMO   LIB('CNF324.DMC.CALPREV.-
LOAD','CNF324.DMC.CALMIC.LOAD',
'CNA324.DMC.TCS')
FORTLIB

WHEN SYSRC(GE 8) END
SCRATCH DEMO.OBJ
CALL DEMO.LOAD(TEMPNAME)
```

AUTOMATED LABORATORY MANAGEMENT SYSTEM IN REGION V

By David Rockwell

INTRODUCTION

The Laboratory Data Management System (LDMS) was developed to process data resulting from compliance monitoring requirements of EPA legislation. Our objective was to complete a compliance monitoring report on each survey within 30-35 days after the completion of the survey field study.

On the average, each compliance monitoring study takes three days. Data are collected from a specific station for a 24-hour period. The types of samples are "grab samples"; twelve grab samples are integrated over a 24-hour period.

In Region V these compliance monitoring studies are performed by four district offices (DO) located in different states. Each DO collects, analyzes, and transmits water samples to the Central Regional Laboratory (CRL) in Chicago.

CHARACTERISTICS OF LDMS

This project provides a beginning stage in the quality control of data going into the Storage and Retrieval (STORET) system. The advantages of LDMS are rapid exchange of analytical data between the CRL and DO. This capability is very useful in emergency situations. It also allows ready reproduction of data values for reports prior to storage of the data in STORET as well as a simpler means for correcting data values. All data from a survey are organized and available in one place at all times during the study period. This data availability assists laboratory managers, chemists, field engineers, and ADP personnel in performing their jobs more effectively. Quality control of data values is provided by means of the SIDES and QUALITY CONTROL programs. Better quality assurance of data handling by knowledgeable DO and CRL personnel is possible, as contrasted with the keypunch approach used previously. With the LDMS, CRL has experienced fewer reanalyses problems (30 to 40 percent fewer samples must be analyzed) in studies because sample descriptions are now always complete. The CRL can better schedule and plan its incoming work load when studies are identified in LDMS.

The disadvantages of the system are the heavy dependence on a central computing facility, the need to train an uninformed user community in a new skill, the use of computer terminals and computers for data entry, and the strain on the organization to run two data handling systems with the same personnel during development and implementation of the LDMS.

Common to all automated systems is the problem of inputting the data. This problem has not been satisfactorily solved in spite of the use of LOAD and GO programs designed to reduce the need for advanced ADP skills.

DATA FLOW

In carrying out a survey, DO engineer plans a study to monitor a discharger. He selects station sites, STORET water quality parameters, and CRL log numbers prior to survey. In LDMS, this information is given to the DO ADP staff and is input into the LABORATORY program.* Figure 1 shows the programs and procedures which comprise LDMS. Figure 2 shows the overall steps in the process of entering data.

LABORATORY is run and produces LABEL. The DO ADP staff enters the LABEL name in the DO LABORATORY DATA CONTROL SYSTEM† (LDCS) data set (see Figure 3). The important columns comprising Figure 3 are numbered, and the columns can be identified as follows:

(1) LABEL data set name.

(2) Disk volume number on Optimum Systems Incorporated where LABEL data set name is found

(3) Compliance monitoring study area description

*Programs LABORATOR, INTERFACE, STORFIT, and the file LABEL were created by Mr. Jon Abraytis, Senior Programmer, Region V, EPA Data Systems Branch, Management Division.

†This LDCS was created by Richard Shekell, Indiana District Office Region V, EPA.

(4) Date when DO completes their data entry

(5) Date when CRL completes their data entry

(6) Date when data are processed through SIDES

(7) Date when data are stored in STORET water quality file

(8) Date when storage of data in STORET is verified, after which LABEL and related data sets are scratched from OSI disks.

This LDCS is used by the CRL to retrieve the various LABEL data sets indicated on Figure 3 in order to plan work schedules for anticipated, incoming field studies.

Each field study is run by the DO field engineer. The DO laboratory analysis of time-dependent parameters (e.g., BOD and phenols), together with field parameters, (e.g., pH and temperature) are entered into LABEL by the DO ADP staff. Figure 4 shows the LABEL data set after the parameter values have been entered. The various parts of this data set are identified as follows:

(1) Number of parameters in LABEL

(2) Number of CRL log numbers assigned to study

(3) STORET agency code of stations used

(4) STORET station types code description

(5) Expected study date

(6) Expected sample arrival date at CRL

(7) Expected analysis due date

(8) Compliance monitoring study area description

(9) CRL log numbers assigned to study

(10) STORET station numbers

(11) STORET sample time and type descriptions in STORET code

(12) Sample site descriptions

(13) STORET parameters grouped in sets of ten

(14) DATA values from study analyses.

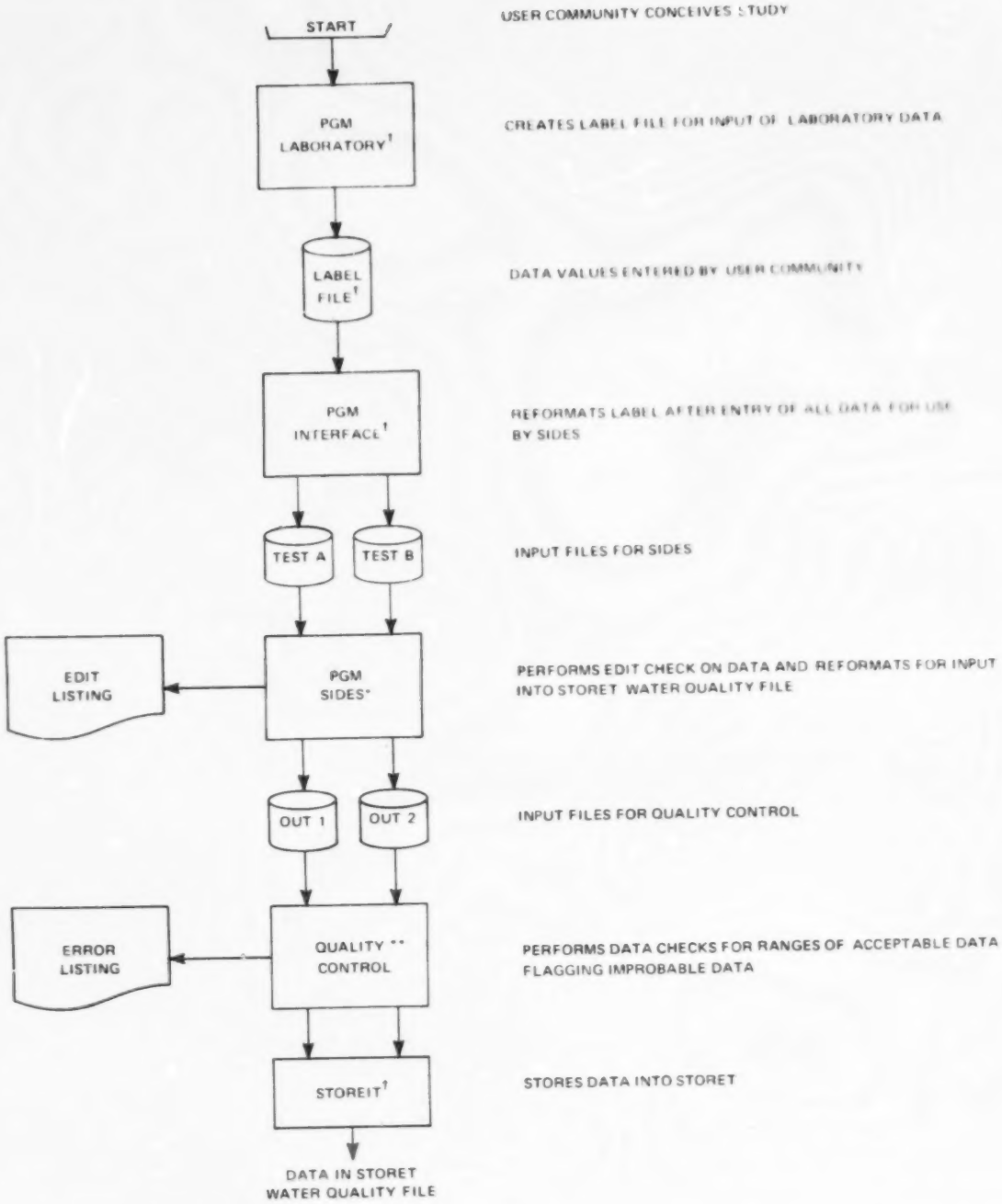
LABEL can handle a maximum of 60 water quality parameters and 190 CRL log numbers. This data set uses a maximum of eight tracks of 3300 disk space. Region V LABEL data sets can usually be stored on two tracks. After the DO ADP staff has entered a date for completion of its work in the LDCS, the major responsibility for inputting data into LABEL is passed to the CRL ADP staff. The CRL analyzes the samples for metals (e.g., mercury and lead), organics (e.g., pesticides), and inorganics (e.g., nutrients). The results of these analyses can be entered on blank copies of the LABEL and given to CRL ADP staff for entry into LABEL. After entry is complete, the CRL ADP staff enters a date in the LDCS under CRL data completion column.

Data in a complete LABEL data set are then ready for the editing and quality assurance programs, SIDES and QUALITY CONTROL (QC)-SCREEN, which are executed consecutively. SIDES flags any format errors to achieve correct sample identification codes and formats (see Figure 2). The QC program eliminates physically impossible data and prevents their being stored into STORET. It prevents physically impossible results, such as a dissolved value exceeding the total value for any parameter. It flags improbable data for the CRL quality assurance officer and DO field engineer to investigate. Positive action must be taken to prevent storage of "improbable" data into STORET.

The LDMS system is replacing a manual system for filling out the Region V data form. Page one of this form (Figure 5) contains study and sample identification information. Page two (Figure 6) is the first of 17 forms for data value entry.

ACKNOWLEDGMENTS

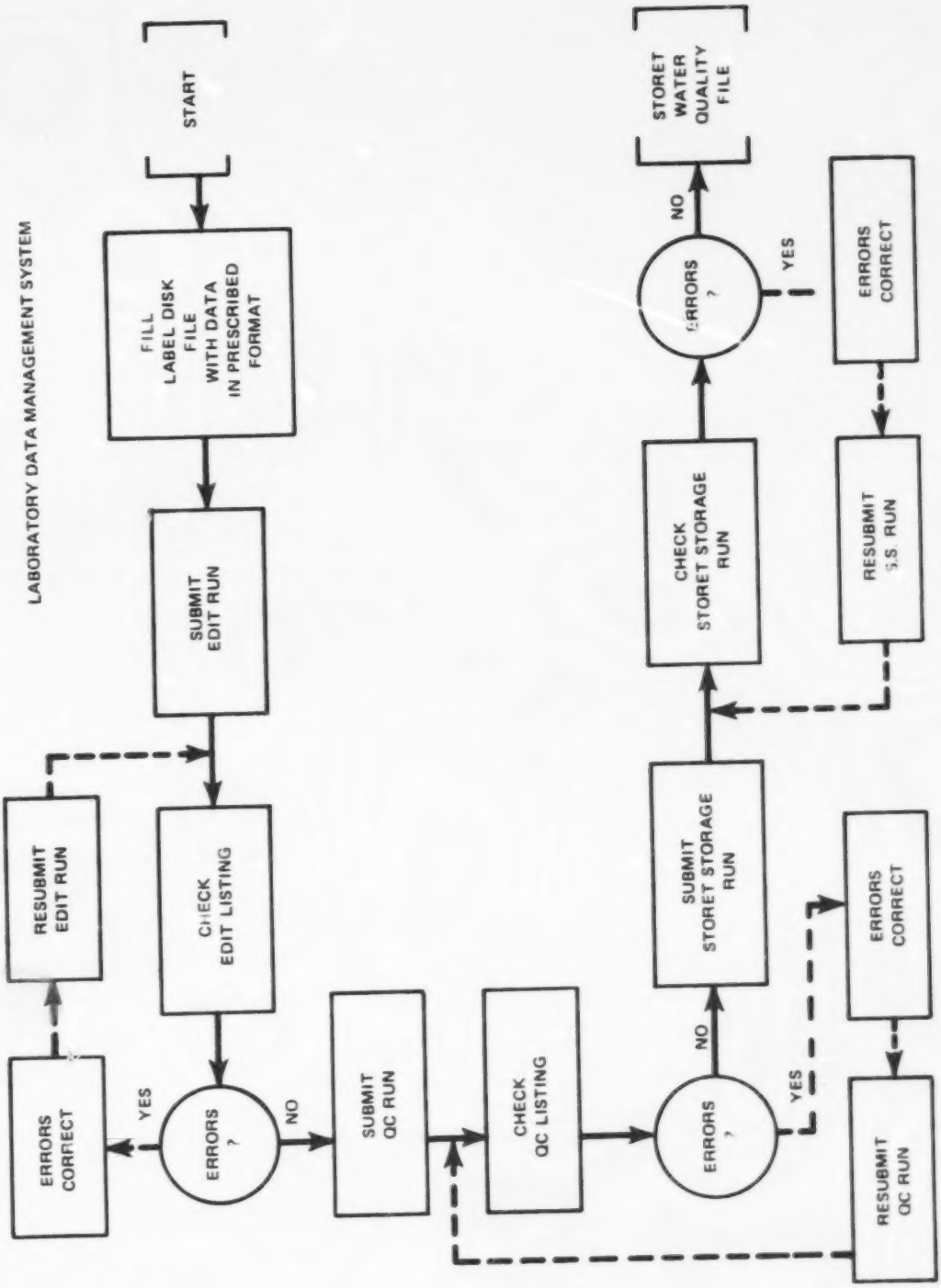
This report is based on the interim results of a development team effort. The number of active persons involved are too numerous to mention; however, the following contributed significantly to the present system: Mr. Jon Abraytis, Mr. Dave Barrow, Dr. Billy Fairless, Mr. Jim Gangler, Dr. Wayne Ott, and Mr. Richard Shekell.



*PROGRAMMED BY DAVE BARROW OF ATHENS, REGION IV, EPA
**PROGRAMMED BY JIM GANGLER, VITRO LABS, INC., WASHINGTON, D.C.
†PROGRAMMED BY JON ABRAYTIS, CHICAGO REGION V, EPA

Figure 1
Laboratory Data Management System
Structure Diagram

Figure 2
System Process Flow Chart



INDIANA DISTRICT OFFICE 111 EAST DIAMOND AVE. EVANSVILLE INDIANA 47711

LISTING OF DATA SET NAMES +VOLUMES +COMMENTS AND DATES

IF THERE ARE ANY QUESTIONS PLEASE CONTACT RICHARD SHERELL

[illegible] FTS OR [illegible] EXT 284

REF. THESE DATA SETS ARE USING ACNA500.DAS.DATA SET NAME OR VOLUME

LAST UPDATE = 740919

(1) LABEL DSN	(2) LABEL VOL	(3) COMMENTS	(4) DATE COMP U.O.	(5) DATE COMP CHL	(6) DSN TESTA VOL	DSN TESTM VOL	DATE ENTM SIDES	DSN OUT1 VOL	DSN OUT2 VOL	(7) DATE STORED	(8) DATE VERIFY
[illegible]	[illegible]	CALCULATED STREAM LOADS	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
[illegible]	[illegible]	CALCULATED [illegible] SAND	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
[illegible]	[illegible]	CALCULATED MUN. LOADS	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
[illegible]	[illegible]	TEST TO EXTENSION DATA SETS	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
[illegible]	[illegible]	THERMAL ELECTRIC STUDY	JUNE	JUNE	[illegible]	[illegible]	740904	[illegible]	N.A.	740904	740914
[illegible]	[illegible]	[illegible] STUDY MUN. TREATMENT PL	JUNE	JUNE	[illegible]	[illegible]	740910	[illegible]	N.A.	740910	740914
[illegible]	[illegible]	[illegible] STUDY TREATED IND. EFF.	JUNE	JUNE	[illegible]	[illegible]	7409	[illegible]	N.A.		
[illegible]	[illegible]	[illegible] STUDY RAW IND. EFF.	JUNE	JUNE	[illegible]	[illegible]					
[illegible]	[illegible]	[illegible] STUDY WATER QUALITY STA.	JUNE	JUNE	[illegible]	[illegible]					
[illegible]	[illegible]	UTILIZER COMPLAINT ANS. AL COMMITT	SCHEDULED								
[illegible]	[illegible]	INFLUENT STATIONS 740522 [illegible]	JUNE	JUNE	[illegible]	[illegible]					
[illegible]	[illegible]	INFLUENT STATIONS 740523 [illegible]	JUNE	JUNE	[illegible]	[illegible]					
[illegible]	[illegible]	[illegible] MUN. AT WAMASH IND.	SCHEDULED								
[illegible]	[illegible]	PRL COMP. [illegible] MEANING DIV.	SCHEDULED								
[illegible]	[illegible]	[illegible] COMP. MAIN OUTFALL	JUNE	JUNE	[illegible]	[illegible]	740904	[illegible]	N.A.	740904	740916
[illegible]	[illegible]	[illegible] COMP. COMP. BELL WATER	JUNE	JUNE	[illegible]	[illegible]	740904	[illegible]	N.A.	740904	740916
[illegible]	[illegible]	INFLUENT STATIONS 740819 [illegible]	SCHEDULED								
[illegible]	[illegible]	INFLUENT STATIONS 740820 [illegible]	JUNE	JUNE	[illegible]	[illegible]					
[illegible]	[illegible]	[illegible] STUDY MUD SERIES	JUNE	JUNE	[illegible]	[illegible]	740910	[illegible]	N.A.	740910	740916
[illegible]	[illegible]	[illegible] WATER STUDY SEGMENT 19	740830								
[illegible]	[illegible]	[illegible] WATER SEG. 19 RIVER STUDY	SCHEDULED								
[illegible]	[illegible]	SEGMENT 21 COMPLIANCE MON.	740830								
[illegible]	[illegible]	SEGMENT 26 COMPLIANCE MON. JAN.	740830								
[illegible]	[illegible]	SEGMENT 19-21 COMP. MON. PESTICIDE	740819								
[illegible]	[illegible]	SEGMENT 26 COMP. MON. PESTICIDES	740830								
[illegible]	[illegible]	OIL SAMPLES PICKED UP BY U.S.C.G.	SCHEDULED								
[illegible]	[illegible]	[illegible] QUIP INVEST. INTAKE	JUNE	JUNE	[illegible]	[illegible]	740910	[illegible]	N.A.	740910	740916
[illegible]	[illegible]	[illegible] QUIP INVEST. MAIN OUTFALL	JUNE	JUNE	[illegible]	[illegible]	740910	[illegible]	N.A.	740910	740916
[illegible]	[illegible]	SEGMENT 63 MUN. SLP									
[illegible]	[illegible]	SEGMENT 63 INDUSTRIAL EFFLUENTS									
[illegible]	[illegible]	SEGMENT 63 INDUSTRIAL INTAKES									
[illegible]	[illegible]	SEGMENT 63 WATER QUALITY									
[illegible]	[illegible]	SEGMENT 51 INDUSTRIAL EFF.									
[illegible]	[illegible]	SEGMENT 51 S.T.P. EFFLUENTS									
[illegible]	[illegible]	SEGMENT 51 S.T.P. INFLUENTS									
[illegible]	[illegible]	SEGMENT 51 INDUSTRIAL INTAKES									
[illegible]	[illegible]	SEGMENT 51 WATER QUALITY									
[illegible]	[illegible]	SEGMENT 66 S.T.P. EFFLUENTS									
[illegible]	[illegible]	SEGMENT 66 S.T.P. INFLUENTS									
[illegible]	[illegible]	SEGMENT 66 INDUSTRIAL EFFLUENTS									
[illegible]	[illegible]	SEGMENT 66 HEADWAT MEANS									

Figure 3
Laboratory Data Control System

```
1.   44   6 12INDO   NOECKER  77777777 740501 740503 740518 4124058842        COSHOCTON STUDY  INDO                    DONE H
2.
3.   743136  COS-1[illegible]  740430 1100 T   12 740501 1000 0500      COSHOCTON STP INFLUENT                          ***IL
4.   743137  COS-1[illegible]  740430 1100 T   12 740501 1000 0000      COSHOCTON STP EFFLUENT                          ***IL
5.   743138  COS-1[illegible]  740501 1000                   0000      COSHOCTON STP EFFLUENT                          ***IL
6.   743169  COS-2[illegible]  740502 1204                   0500      DRESDEN STP INFLUENT                            ***IL
7.   743170  COS-2[illegible]  740501 1000 T   24 740502 1150 0000      DRESDEN STP EFFLUENT                            ***IL
8.   743171  COS-2[illegible]  740502 1150                   0000      DRESDEN STP EFFLUENT                            ***IL
9.
10.  SAMPLE   00[illegible]   00010   00300   00400   00403   00095   00410   00550   31616   00500                  IA
11.  LOG NO   [illegible]   [illegible]   DO      PH      LAB     CNDUCTVY  T ALK   OIL-GRSE  FEC COLI  RESIDUE                IA
12.          [illegible]   [illegible]                   PH      AT 25C    CACO3   TOT-SALT  MFM-FCBR  TOTAL                  IA
13.          [illegible]   [illegible]   MG/L    SU      SU      MICROMHO  MG/L    MG/L      /100ML    MG/L                   IA
14.
15.  743136   |       |       |       |       |7.4    |320.   |160.0  |       |       |710.0                  ***IA
16.  743137   |       |       |       |       |7.1    |455.   |100.0  |       |       |600.0                  ***IA
17.  743138   |[illegible]  |       |       |       |       |       |       |       |K10.0  |                       ***IA
18.  743169   |       |[illegible]  |       |[illegible]  |7.2    |1480.  |340.0  |       |       |1170.0                 ***IA
19.  743170   |       |       |       |       |7.4    |1590.  |310.0  |       |       |985.0                  ***IA
20.  743171   |[illegible]  |[illegible]  |[illegible]  |       |       |       |       |L10000. |                       ***IA
21.
22.  SAMPLE   [illegible]   [illegible]   00[illegible]   00310   00324   00340   00[illegible]   00940   00945   00950                  IB
23.  LOG NO   [illegible]   [illegible]   [illegible]   [illegible]   [illegible]   [illegible]   [illegible]   CHLORIDE  SULFATE  FLUORIDE                IB
24.          [illegible]   [illegible]   [illegible]   [illegible]   [illegible]   HI LEVEL  C       CL      SO4-TOT  F.DISS                  IB
25.          [illegible]   [illegible]   MG/L    MG/L    MG/L    MG/L    MG/L    MG/L    MG/L    MG/L                   IB
26.
27.  743136   |[illegible]  |[illegible]  |[illegible]  |[illegible]  |[illegible]  |350.   |       |84.    |107.   |       |                ***IB
28.  743137   |[illegible]  |[illegible]  |111.   |122.   |[illegible]  |[illegible]  |       |91.    |140.   |       |                ***IB
29.  743138   |       |       |       |       |       |       |       |       |       |       |                       ***IB
30.  743169   |[illegible]  |[illegible]  |[illegible]  |[illegible]  |[illegible]  |       |       |       |       |                ***IB
31.  743170   |[illegible]  |[illegible]  |[illegible]  |[illegible]  |[illegible]  |       |       |210.   |[illegible]  |       |                ***IB
32.  743171   |       |       |       |       |       |       |       |       |       |       |                       ***IB
33.
34.  SAMPLE   [illegible]   [illegible]   00[illegible]   00630   00610   00625   00929   00937   01105   01045                  IC
35.  LOG NO   [illegible]   [illegible]   [illegible]   [illegible]   [illegible]   TOT KJEL  SODIUM   PTSSIUM  ALUMINUM  IRON                   IC
36.          [illegible]   [illegible]   [illegible]   N-TOTAL   TOTAL   N       NA.TOT   K.TOT    AL.TOT   FE.TOT                 IC
37.          [illegible]   [illegible]   MG/L    MG/L    MG/L    MG/L    MG/L    MG/L    UG/L    UG/L                   IC
38.
39.  743136   |       |[illegible]  |[illegible]  |[illegible]  |110.1  |27.0   |       |       |       |       |                ***IC
40.  743137   |       |[illegible]  |[illegible]  |[illegible]  |[illegible]  |111.   |       |       |       |       |                ***IC
41.  743138   |[illegible]  |       |       |       |       |       |       |       |       |       |       |                ***IC
42.  743169   |       |       |[illegible]  |[illegible]  |114.4  |[illegible]  |160.   |       |       |       |                ***IC
43.  743170   |       |       |111.2  |[illegible]  |[illegible]  |[illegible]  |       |       |       |       |                ***IC
44.  743171   |[illegible]  |       |       |       |       |       |       |       |       |       |                       ***IC
45.
46.  SAMPLE   [illegible]   [illegible]   01002   01097   01027   01034   01032   01042   01051   71900                  ID
47.  LOG NO   [illegible]   [illegible]   ARSENIC  ANTIMONY  CADMIUM  CHROMIUM  CHROMIUM  COPPER   LEAD    MERCURY                ID
48.          [illegible]   [illegible]   AS.TOT   SB.TOT   CD.TOT   CR.TOT   HEX-VAL  CU.TOT   PB.TOT   HG.TOTAL               ID
49.          [illegible]   [illegible]   UG/L    UG/L    UG/L    UG/L    UG/L    UG/L    UG/L    UG/L                   ID
50.
51.  743136   |       |       |K2.    |       |K10.   |K30.   |       |50.    |12.    |0.4    |                ***ID
52.  743137   |       |       |K2.    |       |K10.   |K30.   |       |30.    |10.    |0.2    |                ***ID
53.  743138   |       |       |       |       |       |       |       |       |       |       |                       ***ID
54.  743169   |       |K10.   |K5.    |       |K10.   |K30.   |       |80.    |35.    |0.4    |                ***ID
55.  743170   |       |       |K2.    |       |K10.   |K30.   |       |40.    |35.    |0.4    |                ***ID
56.  743171   |       |       |       |       |       |       |       |       |       |       |                       ***ID
57.
58.  SAMPLE   01067   01102   01147   01092                                                                                IE
59.  LOG NO   NICKEL  TIN     SELENIUM  ZINC                                                                                IE
60.          NI.TOTAL SN.TOT  SE.TOT   ZN.TOT                                                                                IE
```

Figure 4
LABEL

Reproduced from best available copy.

Figure 5
Environmental Protection Agency,
Region V, Basic Data Form

FY 1975 District Office _____ Sampling Date _____ Lab. Arrival Date _____ Analysis Due Date _____

Sample Description: River-02111204 Mun-Trt-Eff-03244240 Ind-Trt-Eff-03444240
Lake-02111402 Mun-Eff-Riv-02211204 Ind-Eff-Riv-02411204 Account No. _____ Study _____
Mun-Raw-03212240 Ind-Raw-Inf-03412240

Mark (A) for samples with concentrations above the maximum limits, mark ( ) for concentrations below maximum, mark (D) for District Office analysis

CRL Sample Log Number	Agency No. 8 digits	Station No. 6 digits	Grab Sample Collection Date or Beginning Composite Time	Depth	Type of Composite: =space =time =both blank=grab F=flow	a=average h=maximum l=minimum blank=entire composite	c=continuous g=grab nn=# samples	Ending Composite Sampling Time	Ft. Blank
			Y Y M M D D H H H H						
75-									
75-									
75-									
75-									
75-									
75-									
75-									
75-									
75-									
75-									
75-									
75-									
75-									
75-									
75-									

CRL 8/I/74

FY 1975 District Office _____ Sampling Date _____ Lab. Arrival Date _____ Analysis Due Date _____

Account No. _____ Study _____

Mark (A) for samples with concentrations above the maximum limits, mark (✓) for concentrations below maximum, mark (D) for District Office analysis

Parameter No.		00010	00020	00300	00400	50060	50064	00058			
CHL Sample Log Number	Sample Description	Water Temp	Air Temp	Dissolved Oxygen	Field pH	Residual Chlorine	Free Chlorine	Flow			
Maximum Concentration Limits		---	---	---	---	---	---	---			
Units		C°	C°	mg/l	pH Units	mg/l	mg/l	gal/min			
5											
5											
5											
5											
5											
5											
5											
5											
5											
5											
5											
75											
75											
75											
75											
Totals											

Figure 6
Environmental Protection Agency, Region V, Basic Data Form

AUTOMATIC DATA PROCESSING AND REGIONAL LABORATORY QUALITY ASSURANCE

By Bill Fairless, Ph.D.

The Central Regional Laboratory (CRL), located in Chicago, Illinois and providing Federal laboratory support for the states of Ohio, Indiana, Illinois, Michigan, Wisconsin and Minnesota was formed in February of 1973 with eleven positions. During the past 20 months, the staff has been expanded to 37 permanent, temporary, and intergovernmental personnel act employees. Figure I shows the current organizational chart. The first branch established in our expansion was the Quality Assurance Branch, and the first position filled was Chief of that Branch. A CRL quality assurance program was, therefore, implemented approximately one year ago and has evolved continually since that time. It has resulted in a substantial improvement in the accuracy of the analytical results reported.

It is believed that the Chicago CRL organizational structure formally separating the quality assurance function from all other analytical functions is a major strength. The CRL has a Quality Assurance Branch which reports directly to the director of the laboratory and, therefore, provides an independent evaluation of all sampling, analytical, and data reduction and reporting procedures. The branch reviews all data to be reported, has access to all bench sheets, instrument log books, quality assurance raw data and data summaries, and approves all analytical methods and laboratory techniques.

Cooperative efforts between the Quality Assurance, Biology and Chemistry Branches have resulted in development of the quality assurance data summary which would interest an ADP group as a problem needing an automatic data processing solution. Page 1 of the summary is shown in Figure 2.

The first item specifies a parameter and its STORET number. It was thought that an analytical method could be associated with each STORET number; but since that is not the case, item 2 is included as a reference only. Details of each procedure are given later in the summary. Item 3 is included to differentiate between older, and presumably more reliable, analytical methods and the recently implemented procedures. Item 4 identifies the dates enclosing the time period of data collection for this summary, and item 5 lists all instruments used to collect the data with the location of the instrument log book and the individual responsible for assuring the proper operation of that instrument. Item 6 describes the concentration range which would be used to report data resulting from the procedure referenced in item 2 and other results included in this summary on a routine basis. Data points outside the working concentration range are either not reported, reported with a qualification (do not place in STORET - explain why), or the sample is reanalyzed after appropriate experimental modifications. Item 7 gives a numerical detection limit which is defined on page 3 for each summary. Item 8 describes the control limits used to flag all suspect data, and item 9 presents the precision. Item 10 is included only after analysis of reference samples that have been prepared by someone outside the administrative group completing the summary. Most often these samples are obtained from Methods Development and Quality Assurance Research Laboratory (MDQARL) or Research Triangle Park (RTP). Items 11 and 12 are self-explanatory.

Page 2 of the summary (Figure 3) is a tabular summary of the data used to arrive at answers for the items described previously and listed on page 1 of the form. The usual statistical parameters are included on the left and the chemist is required to fill in the top (starred) of each column. A slightly different form (Figure 4) is used for duplicates.

A quality assurance data summary is prepared for each unique analytical measurement being made at the CRL in which sufficient data are available. Figure 5 is a typical example describing the analysis of cadmium by flame atomic absorption spectroscopy. Figure 6 shows the supporting data for cadmium, and Figure 7 is an example of the terms that should be defined in order to use the numerical data presented in Figures 6 and 7.

Figures 8, 9 and 10 show data for a "high level" nitrate-nitrite analysis while Figures 11, 12 and 13 pertain to a "low level" procedure. The CRL staff has also designed equipment and procedures to analyze for even lower concentrations of nitrate in the samples collected from the Great Lakes.

The staff performs a considerable number of arithmetical calculations on finished data; still more calculations are required for raw data. There has not been sufficient time to summarize such data as absorbance

readings, extinction coefficients, instrument gain settings, and so forth even though these variables are frequently used by CRL analysts to evaluate quality assurance data.

Hopefully these examples clearly illustrate the kinds of internal quality assurance data that are now collected and evaluated by hand. The data management file of the CRL minicomputer system will include programs and data files to accomplish all of the above automatically. The system is described in greater detail by Roman Bystroff in another paper of these Proceedings. It is expected that once these programs are operational, better estimates of data quality will be provided. Such analyses will be made available quickly to other regional laboratories.

These kinds of efforts by the regional laboratories place increasing responsibilities on the managers of the national data bases to ensure the quality of all the data being placed in each data base. It is questionable that a continued effort to produce more accurate measurements is valuable, if those measurements are only to be lost in a data base containing values of all qualities. In fact, it is predictable that many users will turn away from the existing national data storage bases unless the quality or integrity of the data is assured.

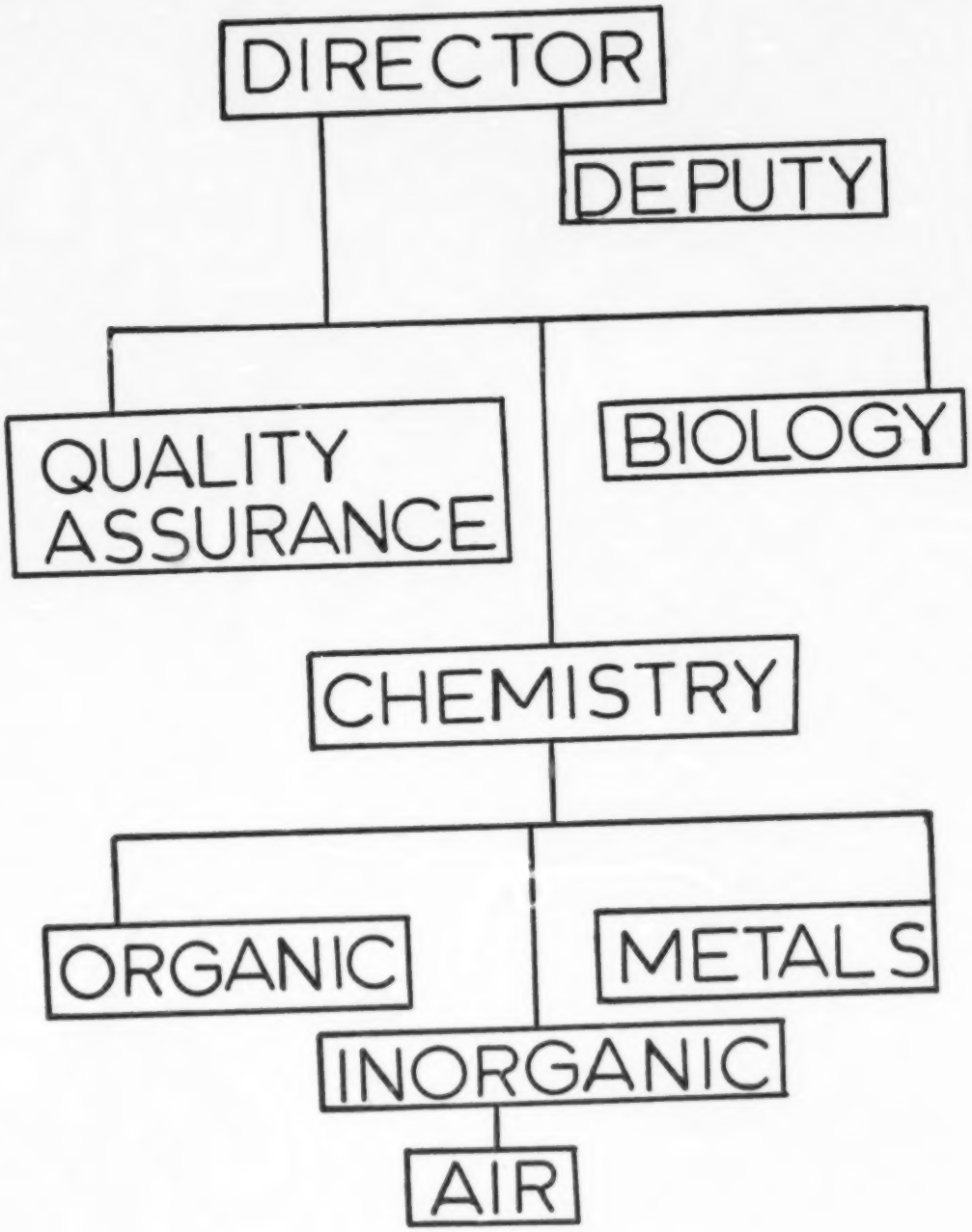


Figure 1
CRL Organizational Chart

REGION V
U.S. ENVIRONMENTAL PROTECTION AGENCY
CENTRAL REGIONAL LABORATORY

Page 1

QUALITY ASSURANCE DATA SUMMARY

1. Parameter _____ STORET No. _____

2. Procedure used _____

_____

_____ Attachment Included /_/ Yes /_/ No

3. Procedure has been in use since _____

4. Data included in this summary was collected from _____ to _____

5. Instruments used to collect data — Location of instrument Log book & individual responsible for instrument maintenance

   a. _____

   b. _____

   c. _____

6. Concentration range reported _____ to _____ units

7. Detection Limit _____ units

8. Control Limits — Percent of data outside control limits

   a. _____

   b. _____

   c. _____

   d. _____

   e. _____

   f. _____

9. The precision is _____ at a concentration of _____ using _____ standard deviations as an estimate of precision.

10. Bias _____

11. Significant figures reported including correct STORET units _____

12. Signature _____ Date _____

    Immediate Supervisor _____ Date _____

Figure 2
Quality Assurance Data Summary

STATISTICAL SUMMARY OF
QUALITY ASSURANCE DATA

	*	*	*	*
1. No. Samples				
2. No. Determn/Sple.				
3. True Values				
4. Mean Value Determn.				
5. Bias				
6. Variance				
7. Average of differences				
8. Std. Deviation of differences				
9. Est. Std. Deviation of Values				
10. 95% Confidence Range				
11. Relative Std. Dev.				
12. 95% Rel. Std. Dev. Range				
13. Data Collected from to				

(a) *Blank - All reagents except sample

(b) Reference standard - An unknown supplied by an external audit agency such as MDQARL

(c) Control standard - Sample of known concentration prepared by CRL staff.

(d) Calibration standard - Sample used to adjust instruments.

(e) Spike - A measured amount of material added to a sample prior to the analysis.

(f) Other _____

_____

_____

Figure 3
Statistical Summary of
Quality Assurance Data

STATISTICAL SUMMARY OF
QUALITY ASSURANCE DATA
FOR DUPLICATE DETERMINATIONS

1. No. Samples				
2. No. Determn/Sple.				
3. True Values				
4. Mean Value Determn.				
5. Bias				
6. Variance				
7. Average of differences				
8. Std. Deviation of differences				
9. Est. Std. Deviation of Values				
10. 95% Confidence Range				
11. Relative Std. Dev.				
12. 95% Rel. Std. Dev. Range				
13. Data Collected from to				

*Blank - All reagents except sample

Standard (a) reference - An unknown supplied by an external audit agency such as MDQARL

(b) control - Sample of known concentration prepared by CRL staff.

(c) calibration - Sample used to adjust instruments.

Spike - Amount of material added to a sample.

Other _____

_____

_____

Figure 4
Statistical Summary of Quality Assurance Data for Duplicate Determination

REGION V
U.S. ENVIRONMENTAL PROTECTION AGENCY
CENTRAL REGIONAL LABORATORY

QUALITY ASSURANCE DATA SUMMARY

1. Parameter tot. Cd STORET No. 01027

2. Procedure used Acid digestion and flame atomic absorption.

3. Procedure has been in use since February 1973

4. Data included in this summary was collected from 7/73 to 5/74

5. All instruments used to collect data — Location of instrument Log book & responsible individual

a.	PE 306	R. Whitworth/Metals
b.	IL 453	E. King/Metals
c.		
d.		
e.		

6. Working concentration range 10 ug/l to 500 ug/l

7. Detection Limit 5 ug/l

8. Control Limits — Percent of data outside control limits

a.		
b.	91-109% Spike recovery	
c.	<10% Sample conc. in blank	or
d.	<10 PPb Cd	

9. The precision is 9 ug/l at a concentration of 100 ug/l using 2 standard deviations as the measure.

10. Bias

11. Units including correct No. Significant Figures 2 or 3 ug/l

12. Signature _____ Date _____

Immediate Supervisor Ed Huff Date 9/20/74

Figure 5
Quality Assurance Data Summary-Analysis of Cadmium by Flame Atomic Absorption Spectroscopy

STATISTICAL SUMMARY OF
QUALITY ASSURANCE DATA
FOR UNIVARIABLES

Cd

	* Spike	* Reference	* Reference	*
1. No. Samples	111	1	1	
2. No. Detl Sample	1	1	1	
3. True Value	100%	16 ug/l	73 ug/l	
4. Mean Value of Determinations	99%	13 ug/l	68 uq/l	
5. Bias	-1%	-3 ug/l	-5 uq/l	
6. Variance	19.4			
7. Standard Deviation	4.4			
8. 95% Confidence Range	91-109			
9. Relative Std. Deviation	4.4			
10. 95% Rel. Confidence Range	91-109			
11. Data Collected from to	7/73 - 5/74			

*Blank - All reagents except sample

*Standard (a) reference - An unknown supplied by an external audit agency such as MDQARL.

(b) control - Sample of known concentration prepared by CRL staff.

(c) calibration - Sample used to adjust instruments.

*Spike - Amount of material added to a sample.

*Other _____

_____

_____

Figure 6
Supporting Data for Cadmium

DEFINITIONS AND DETAILED DESCRIPTIONS
OF TERMS USED ON PREVIOUS PAGES

Include the following

I. Test procedures, working concentration range, detection limit, control limits, precision, accuracy or bias, actions usually taken when data is outside control limits.

1. The test procedure is the standard EPA approved procedure given in "Methods for chemical Analysis of Water and Wastes." U.S. EPA, 1971, page 83-120.

2. The working standards used for instrument calibration are obtained from commercial sources and diluted to the desired concentration with distilled-deionized water.

3. The working concentration range is established by the calibration standards. It begins at a concentration above the detection limit and covers the remaining linear or near linear portion of the absorbance vs concentration curve.

4. The detection limit is defined as that concentration of metal that gives a signal peak twice as large as the baseline noise. It is evaluated from standards that are approximately ten times the detection limit using the equation:

$$\text{Detection Limit} = 2\left(\begin{matrix}\text{Standard deviation}\\ \text{of baseline noise}\end{matrix}\right)\left(\frac{\text{concentration}}{\text{instrument readout}}\right)$$

5. Accuracy is estimated from the long term percent recoveries of spikes added to different kinds of samples as follows:

$$\%\ \text{Recovery} = \frac{[\text{Sample + Spike}] - [\text{Sample}]}{[\text{Spike}]} \times 100$$

If the spike concentration is less than one half the sample concentration the results are not used for control purposes.

6. The apparent or real concentration of reagent blanks are measured and subtracted from the respective samples. If a blank is contaminated so it contains more than 10% of the sample concentrations the analyses are considered to be out of control.

7. Significant figures are determined by the bench chemist and are based on the day to day performance of the laboratory equipment.

8. All analyses that do not fall within the control limits are investigated on an individual basis. Usually the entire series of samples are reanalyzed. In all cases the action taken is recorded in the bench log book. All reported data are supported by Quality Assurance results within the defined control limits or which have been approved by the CRL Quality Assuranch Branch.

Figure 7
Definitions and Detailed Description of Terms

REGION V
U.S. ENVIRONMENTAL PROTECTION AGENCY
CENTRAL REGIONAL LABORATORY

QUALITY ASSURANCE DATA SUMMARY

1. Parameter NO_2+NO_3-N high level STORET No. 00630

2. Procedure used Technicon Industrial Method 158-71 modified to account for H_2SO_4 preservation

3. Procedure has been in use since 7/1/73

4. Data included in this summary was collected from 7/1/73 to 8/12/74

5. All instruments used to collect data — Location of instrument Log book & responsible individual

a. Technicon AA II	Nutrient Lab. - A. Jirka
b. Fisher Diluter	" " " "
c. Balance #104331	Nutrient Lab. - M. Carter
d.	
e.	

6. Working concentration range 0.03 to 5.00

7. Detection Limit 0.03

8. Control Limits 3 std. dev. — Percent of data outside control limits

a. Blk: 0.00 - 0.04	3%
b. 1.00: 0.93 - 1.11	1%
c. 3.00: 2.85 - 3.09	12%
d. 5.00: 4.84 - 5.14	9%
recovery 90-108%	2%

9. The precision is + 0.12 at a concentration of 3.00 mg/l using 2 standard deviations as the measure.

10. Bias -0.01 average

11. Units including correct No. Significant Figures X.XX MG/L

12. Signature Andrea Jirka Date 9/4/74

Immediate Supervisor Mark J. Carter Date 9/18/74

Figure 8
Data Summary, High-Level Nitrate-Nitrite Analysis

STATISTICAL SUMMARY OF
QUALITY ASSURANCE DATA
FOR UNIVARIABLES

	*Control Std. Blank	* C. Std. 1.00 mg/l	* C. Std. 3.00 mg/l	* C. Std. 5.00 mg/l
1. No. Samples	58	75	40	57
2. No. Detl Sample	1	1	1	1
3. True Value	0.00	1.00	3.00	5.00
4. Mean Value of Determinations	0.01	1.02	2.97	4.98
5. Bias	+0.01	+0.02	-0.03	-0.02
6. Variance	0.00015	0.00096	0.0037	0.0051
7. Standard Deviation	0.012	0.031	0.061	0.071
8. 95% Confidence Range	0-0.003	0.95-1.08	2.85-309	4.84-5.12
9. Relative Std. Deviation	——	3.1%	3.0%	1.4%
10. 95% Rel. Confidence Range	——	95 - 108%	95 - 103%	97 - 102%
11. Data Collected from to	7/1/73 to 8/13/74	7/1/73 to 8/13/74	1/11/74 to 8/13/74	7/1/74 - 8/13/74

*Blank - All reagents except sample

Standard (a) reference - An unknown supplied by an external audit agency such as MDQARL.

(b) control - Sample of known concentration prepared by CRL staff.

(c) calibration - Sample used to adjust instruments.

* Spike - Amount of material added to a sample.

* Other _____

_____

_____

Figure 9
Data Summary, High-Level
Nitrate-Nitrite Analysis

STATISTICAL SUMMARY OF
QUALITY ASSURANCE DATA
FOR UNIVARIABLES

	* C. Std. 2.5	* Spikes	* Std. Cal.	* Duplicates
1. No. Samples	25	47	26	22
2. No. Detl Sample	1	1	1	2
3. True Value	2.5	100%	—	—
4. Mean Value of Determinations	2.48	99%	3.40	—
5. Bias	-0.02	-1%	C	Ave of
6. Variance	0.0014	9.9%	O	difference
7. Standard Deviation	0.038	—	N T R	is 0.01 &
8. 95% Confidence Range	2.40 to 2.56	—	O L =	Std. dev. is
9. Relative Std. Deviation	1.5%	3.1%	3.40 ± 1	0.016
10. 95% Rel. Confidence Range	96 to 102%	93 - 105%		
11. Data Collected from to	7/1/73 1/11/74	7/1/73 8/13/74		

*Blank - All reagents except sample

Standard (a) reference - An unknown supplied by an external audit agency such as MDQARL.

(b) control - Sample of known concentration prepared by CRL staff.

(c) calibration - Sample used to adjust instruments.

* Spike - Amount of material added to a sample.

* Other _____

_____

_____

Figure 10
Data Summary, High-Level
Nitrate-Nitrite Analysis

REGION V
U.S. ENVIRONMENTAL PROTECTION AGENCY
CENTRAL REGIONAL LABORATORY

QUALITY ASSURANCE DATA SUMMARY

(low level)

1. Parameter $(NO_2 + NO_3)$-N STORET No. 00630

2. Procedure used Technicon Industrial Method No. 158-71W with or without modification to account for H_2SO_4 preservation.

3. Procedure has been in use since 7/1/74

4. Data included in this summary was collected from 7/1/74 to 8/13/74

5. All instruments used to collect data — Location of instrument Log book & responsible individual

   a. Technicon AA-II — Nutrient Lab. - A. Jirka

   b. Balance #104331 — Nutrient Lab. - M. Carter

   c. ____ — ____

   d. ____ — ____

   e. ____ — ____

6. Working concentration range 0.002 to 1.000

7. Detection Limit 0.002

8. Control Limits 3 std. dev. — Percent of data outside control limits

   a. Blk: 0.000 - 0.003 — 0

   b. 0.2: 0.195 - 0.205 — 0

   c. 1.0: 0.979 - 1.013 — 10%

   d. % Rec: 88% - 112% — 0

9. The precision is ± 0.003 at a concentration of 0.200 using 2 standard deviations as the measure.

10. Bias -0.002 average

11. Units including correct No. Significant Figures 0.XXX mg/l

12. Signature Andrea Jirka Date 9/5/74

    Immediate Supervisor Mark J. Carter Date 9/18/74

Figure 11
Data Summary, Low-Level
Nitrate-Nitrite Analysis

STATISTICAL SUMMARY OF
QUALITY ASSURANCE DATA
FOR UNIVARIABLES

	*c Blk	*c 0.200	*c 0.500	*c 0.600
1. No. Samples	18	13	6	8
2. No. Detl Sample	1	1	1	1
3. True Value	0.000	0.200	0.500	0.600
4. Mean Value of Determinations	0.001	0.200	0.500	0.593
5. Bias	0.001	0.000	0.000	-0.007
6. Variance	0.00000047	0.0000026	1.0000002	0.000038
7. Standard Deviation	0.00068	0.0016	0.00048	0.0062
8. 95% Confidence Range	0.000-0.002	0.197-0.203	0.499-0.500	0.581-0.605
9. Relative Std. Deviation	——	0.8%	0.1%	1%
10. 95% Rel. Confidence Range	——	98.5-101.5%	99.8 - 100.2%	97 - 101%
11. Data Collected from to	7/1/74 9/5/74	7/1/74 9/5/74	7/1/74 9/5/74	7/1/74 9/5/74

*Blank - All reagents except sample

Standard (a) reference - An unknown supplied by an external audit agency such as MDQARL.

(b) control - Sample of known concentration prepared by CRL staff.

(c) calibration - Sample used to adjust instruments.

*Spike - Amount of material added to a sample.

*Other _____

_____

_____

Figure 12
Data Summary, Low-Level
Nitrate-Nitrite Analysis

STATISTICAL SUMMARY OF
QUALITY ASSURANCE DATA
FOR DUPLICATE DETERMINATIONS

	c 1.000	Std. Cal.	% Rec.	Duplicates
1. No. Samples	10	2	8	31
2. No. Determn/Sple.	1	1	1	2
3. True Values	1.000	——	100%	
4. Mean Value Determn.	0.996	3.20	100%	
5. Bias	-0.004		0	
6. Variance	0.000031		15	
7. Average of differences	——		——	0.002
8. Std. Deviation of differences	——		——	0.0026
9. Est. Std. Deviation of Values	0.0056		——	
10. 95% Confidence Range	0.985-1.007		——	
11. Relative Std. Dev.	0.6%		3.9%	
12. 95% Rel. Std. Dev. Range	98-101%		92-108%	
13. Data Collected from to	7/1/74 9/5/74		7/1/74 9/5/74	

*Blank - All reagents except sample

Standard (a) reference - An unknown supplied by an external audit agency such as MDQARL

(b) control - Sample of known concentration prepared by CRL staff.

(c) calibration - Sample used to adjust instruments.

Spike - Amount of material added to a sample.

Other _____

_____

_____

Figure 13
Data Summary, Low-Level
Nitrate-Nitrite Analysis

SAMPLE TRACKING DATA MANAGEMENT SYSTEM

By G. C. Allison, M. J. Madsen, and R. N. Snelling*

INTRODUCTION

A Sample Tracking Data Management System (STDMS) is under development at the National Environmental Research Center in Las Vegas, Nevada (NERC-LV) to provide a capability for laboratory control (sample tracking) as well as the management of environmental surveillance data.

The NERC-LV is involved in a variety of environmental surveillance projects. In support of these efforts, Laboratory Operations receives on the order of 3,000 samples per month for specific stable chemical and/or radiochemical analysis. Sample types include air filters, gas samples, milk, water, edible and nonedible vegetation, animal tissue, and soil. The specific analyses performed vary with sample type and project objectives. Table I depicts the major analysis "families" along with typical monthly sample loads and expected processing times. Anywhere from 1 to 27 specific analyses may be required for a given sample. Expected processing times range from 1 to 16 weeks. The indicated sample load coupled with the wide range of analyses and processing times results in approximately 7,500 sample-analysis pairs in the system at any point in time.

SAMPLE AND DATA FLOW

The primary flow of data and samples within the laboratory is depicted in Figure 1. This schematic is typical of an organization not having ADP capability. Four functional activities are identified:

- Project Management: the planning, direction, and coordination of project activities and the ultimate analysis and interpretation of data.
- Field Operations: the collection of physical samples and *in situ* data.
- Sample Control: the receiving, recording, and routing of incoming samples and data.
- Laboratory Operations: the stable chemical and radiochemical analysis of samples and the subsequent generation of parametric data.

The progression of these activities is as follows. Project Management establishes the protocol for Field Operations and Laboratory Operations sample analysis. Field Operations collects samples and data and delivers them to Sample Control. Sample Control serves essentially as the central receiving and accounting point for all incoming data and samples. The samples are routed to Laboratory Operations for analysis, with analysis results being reported back to Project Management for review and interpretation.

AUTOMATIC DATA PROCESSING FUNCTION

The automatic data processing function is utilized to integrate all these elements. A system which has been operational at the NERC-LV for several years is shown conceptually in Figure 2. The system provides basically for the storage and retrieval of data from a direct access master file. On the input side, sample identification and collection data are entered through Sample Control through the use of an "H" card (sample header information) and an "L" card (location description information). A unique six-digit sample number is assigned to each physical sample and is used to identify all data records for that sample.

A unique part of the system is that the laboratory analytical programs are interfaced directly with the master file for the purpose of retrieving needed sample collection information (such as collection date for use in radioactive decay calculations). Analytical result cards "R" cards) are generated for direct reentry to the data management system after review by Laboratory Operations.

On the output side of the system, data summaries may be generated directly from the master file. A typical report is shown in Figure 3. In addition, working files may be created for the purpose of utilizing special purpose application programs such as dose models, plotting, contouring, and statistical analysis. The weakness of the system as shown in Figure 2 is that there is no mechanism for ascertaining the analysis status of a given sample other than its presence or absence in the master file. With an average of 7,500 analyses in various stages of completion at any point in time, this does not provide

*Speaker

Table I
NERC-LV Analysis Distribution, 1973

Analysis Family	Specific Analysis	Sample Load (Number per Month)	Expected Processing Time (Days)
Radiochemistry			
Gamma (NaI)	All present	350	7
Gamma (GeLi)	All present	10	14
Gross α, β	--	120	21
Strontium	^{89}Sr, ^{90}Sr	50	45
Tritium	^{3}H	165	21
Tritium (enriched)	^{3}H	20	60
Carbon	^{14}C	1	90
Radium	^{226}Ra, ^{228}Ra	23	60
Iron	^{55}Fe	1	60
Plutonium	^{238}Pu, ^{239}Pu	50	90
Uranium	^{234}U, ^{235}U, ^{238}U	32	90
Thorium	^{228}Th, ^{230}Th, ^{232}Th	5	90
Polonium	^{210}Po	4	90
Americium	^{241}Am	1	90
Noble gases	Kr, Xe, Ar	50	30
Radon	^{222}Rn	8	7
Iodine	^{129}I	1	30
Lead	^{210}Pb	1	120
Stable Chemistry			
Ash	Ash Weight	2	45
Major constituent - Water	Ca, Sr, Li, Na, K, Mn, Mg, Al, Fe, Si, B, F, Cl, SO_4	6	21
Nutrient Analysis - Water	Alk, P-diss, P-tot, NH_3-N, NO_2 + NO_3 - N, TKN	820	21
Selenium	Se	1	21
Cadmium	Cd	1	21
Arsenic	As	1	21
Molybdenum	Mo	1	21
Vanadium	V	1	21
Mercury	Hg	1	21
Lead	Pb	1	21
Zinc	Zn	1	21
Thorium	Th	1	21

the degree of control or depth of information required for the management of Laboratory Operations.

SAMPLE TRACKING

The sample tracking module is now being implemented to provide a tool to management for evaluating and controlling the Laboratory Operations function by providing sample status information. This system is shown in Figure 4. A status file has been added which contains status statistics. Upon sample log-in, analysis to be performed is defined either by default or through the use of an "M" card. As completed analyses are received ("R" cards), the flags are turned off. A variety of status reports may be obtained by an exception reporting technique. These reports are utilized by both laboratory and project management for assessing sample status.

STDMS SYSTEM DESCRIPTION

Figure 5 shows a more detailed schematic of the STDMS. The system is organized around two files, the master file, containing all parametric and collection data for each sample number, and the status file, containing status related data for each sample number. A third file, the location description file, is used to retrieve standard narrative location descriptions and coordinates. Working files may also be created for special applications, such as statistical analysis, exposure modeling, and interfacing with external data bases (STORET).

The system enters and updates data using four different card types (Figure 6). The "H" card is used to initiate a sample number and generate a data record on the master file and status file with each new sample. Analyses to be performed and expected analysis times are established by default according to project code and sample type code. The analyses have been combined into related groups called families.

The default family parameters may be overridden through the use of the "M" card. The "M" card is used to update the families requested and modify the report due date for a given sample number on the status file.

For samples which were collected at locations not on the location description file, an "L" card is used to provide narrative location description information along with the station coordinates. Card type "R" is used to add results from a specific analysis family to the master file and to update the status file by turning off the flag for the completed analysis family.

The master file and status files are index sequential files designed to permit direct access through the use of multiple keys. The keys, defined at file creation time, allow retrieval to be made according to user reporting needs. The key definition generates levels of indices which contain pointers to the appropriate data records. Sample number, project code, and sample type are the primary data elements utilized to form a key.

An array of constants is maintained defining the expected analysis time for each family of analyses. Expected analysis time is defined as the amount of time in days required to complete the longest analysis within a family. When a status report is generated, the expected analysis time is added to the log-in date and the resulting date is then compared to the current date to determine an overdue analysis. The calculated analysis due date may be overridden with the report due date specified on the "M" card. Thus the report due date can be set prior to the calculated analysis due date.

STATUS REPORTS

The analysis in process report (Figure 7) lists only those samples with outstanding analyses to be completed. Samples are listed in order by sample number, within project, within analysis family. Samples which have exceeded their expected completion date are indicated with an asterisk. This format is intended for use by the laboratory analysis personnel. The analysis in process by program report lists samples which have analyses to be performed within a project (Figure 8). Past due samples are flagged. This report is designed for middle management use. The status summary by project report (Figure 9) is intended for project management. It summarizes samples received, analyses requested, and status. These reports allow the STDMS to provide NERC-LV management the capability of tracking all samples through their respective analyses to completion.

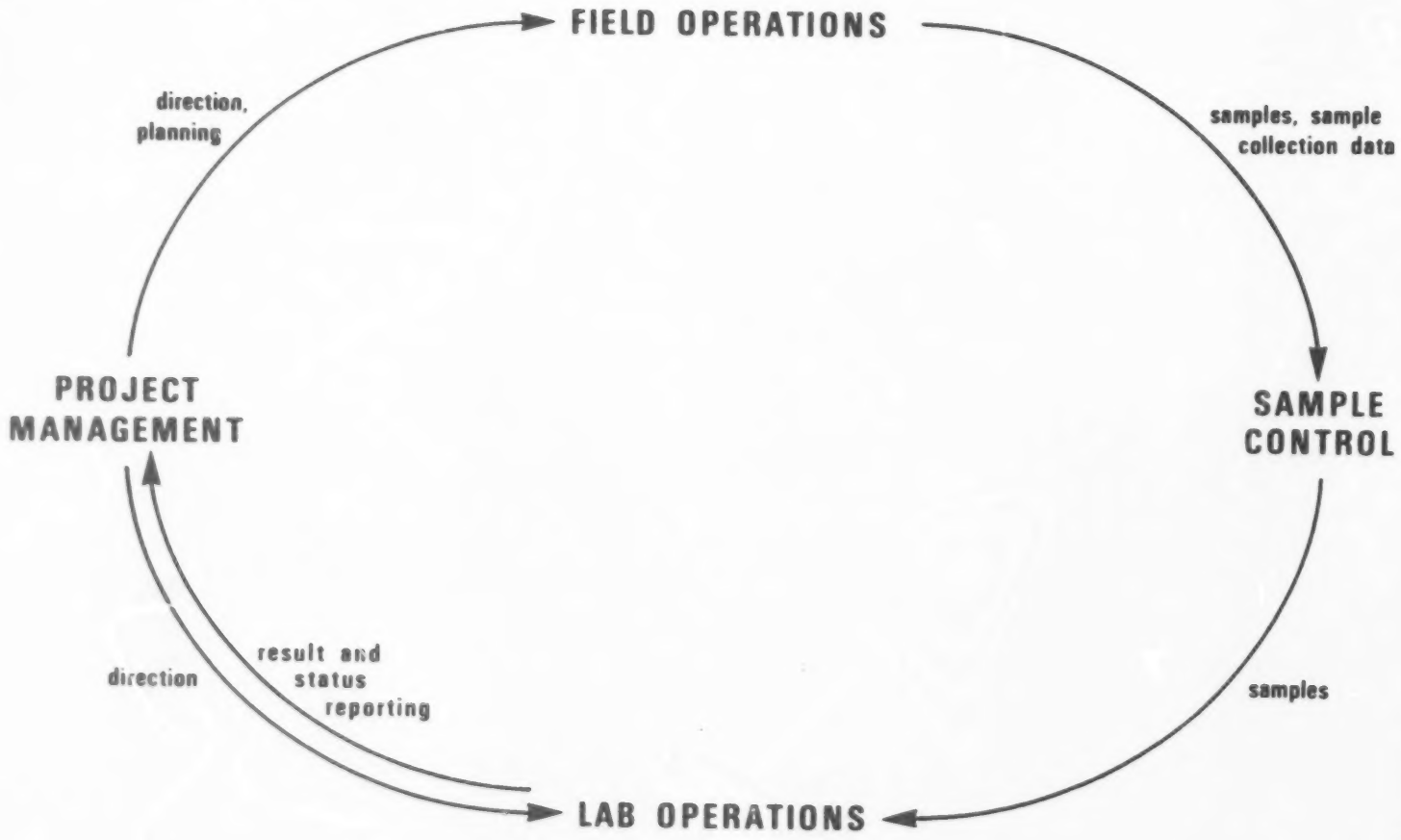
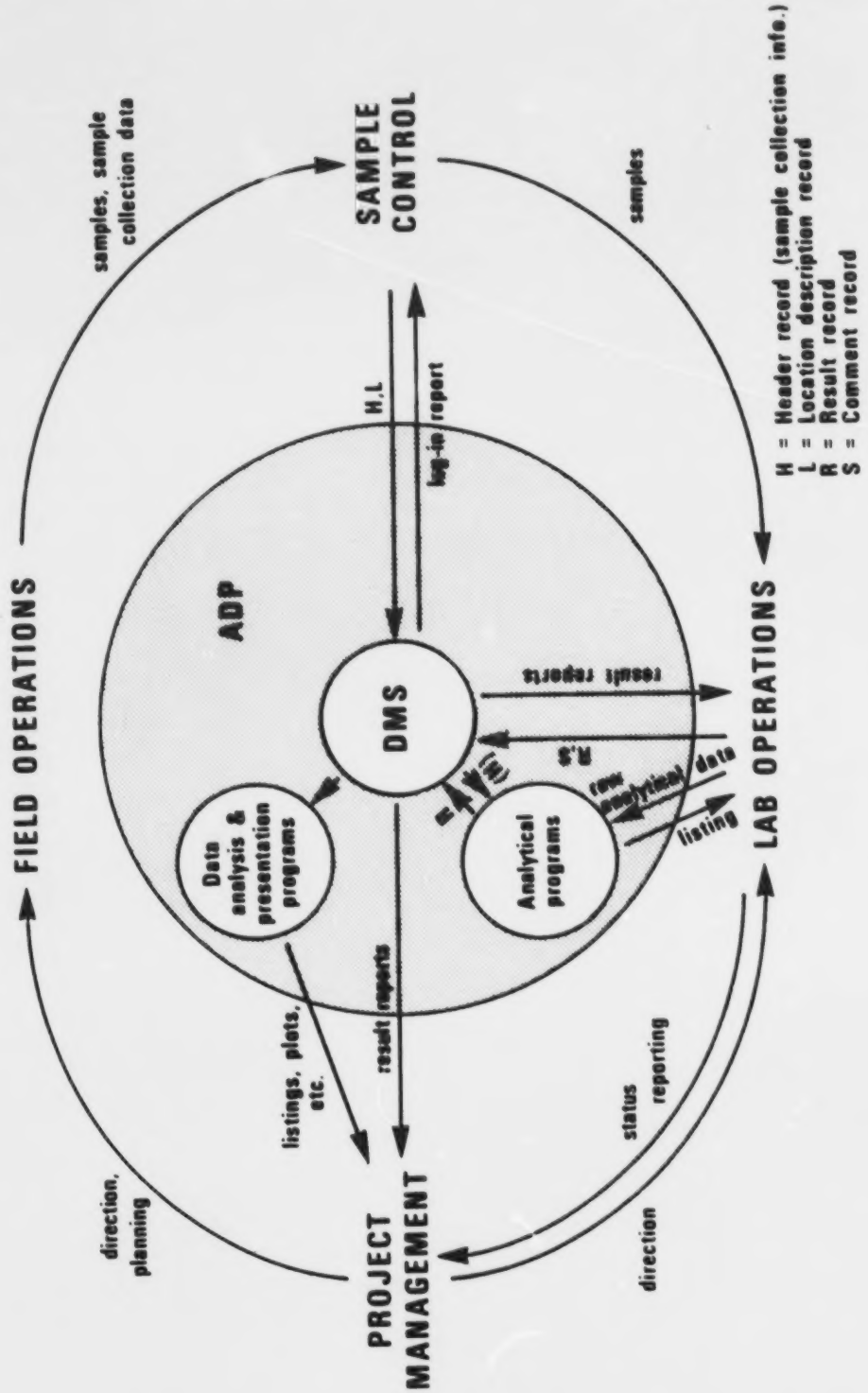


Figure 1
Primary Sample and Data Flow

Figure 2
Automatic Data Processing Function



SAMPLE REPORT (HYPOTHETICAL DATA) PAGE 001

WASHINGTON REPORTED 09/10/71	90SR PCI/L	131I PCI/L	132TE PCI/L	137CS PCI/L
CHEHALIS WASH - CONSOLIDATED DAIRY 000001 11 0005 DATE- 06 30 71 1200 SIZE- 3.50 L	NA	2.2E02 (4.4E01)	7.5E02 (1.4E01)	6.1E01 (1.0E01)
CHEHALIS WASH - CONSOLIDATED DAIRY 000002 11 0005 DATE- 06 30 71 1200 SIZE- .400 L	4.2E02 (1.5E01)	LT3E01	3.9E01 (2.7E01)	4.8E03 (1.8E02)
CHEHALIS WASH - CONSOLIDATED DAIRY 000003 11 0005 DATE- 06 30 71 1200 SIZE- 3.50 L	LT3E01	7.6E02 (3.9E01)	4.7E01 (1.9E01)	8.6E01 (1.6E01)
CHEHALIS WASH - CONSOLIDATED DAIRY 000004 11 0005 DATE- 06 30 71 1200 SIZE- .400 L	LT3E01	1.0E01 (1.0E01)	1.1E01 (1.0E01)	7.9E02 (4.3E01)
CHEHALIS WASH - CONSOLIDATED DAIRY 000005 11 0005 DATE- 06 30 71 1200 SIZE- .400 L	LT4E01	ND	ND	ND
CHEHALIS WASH - CONSOLIDATED DAIRY 000006 11 0005 DATE- 06 30 71 1200 SIZE- .400 L	NA	ND	ND	ND

Figure 3
Sample Report (Hypothetical Data)

Figure 4
Data Management System with Sample Tracking

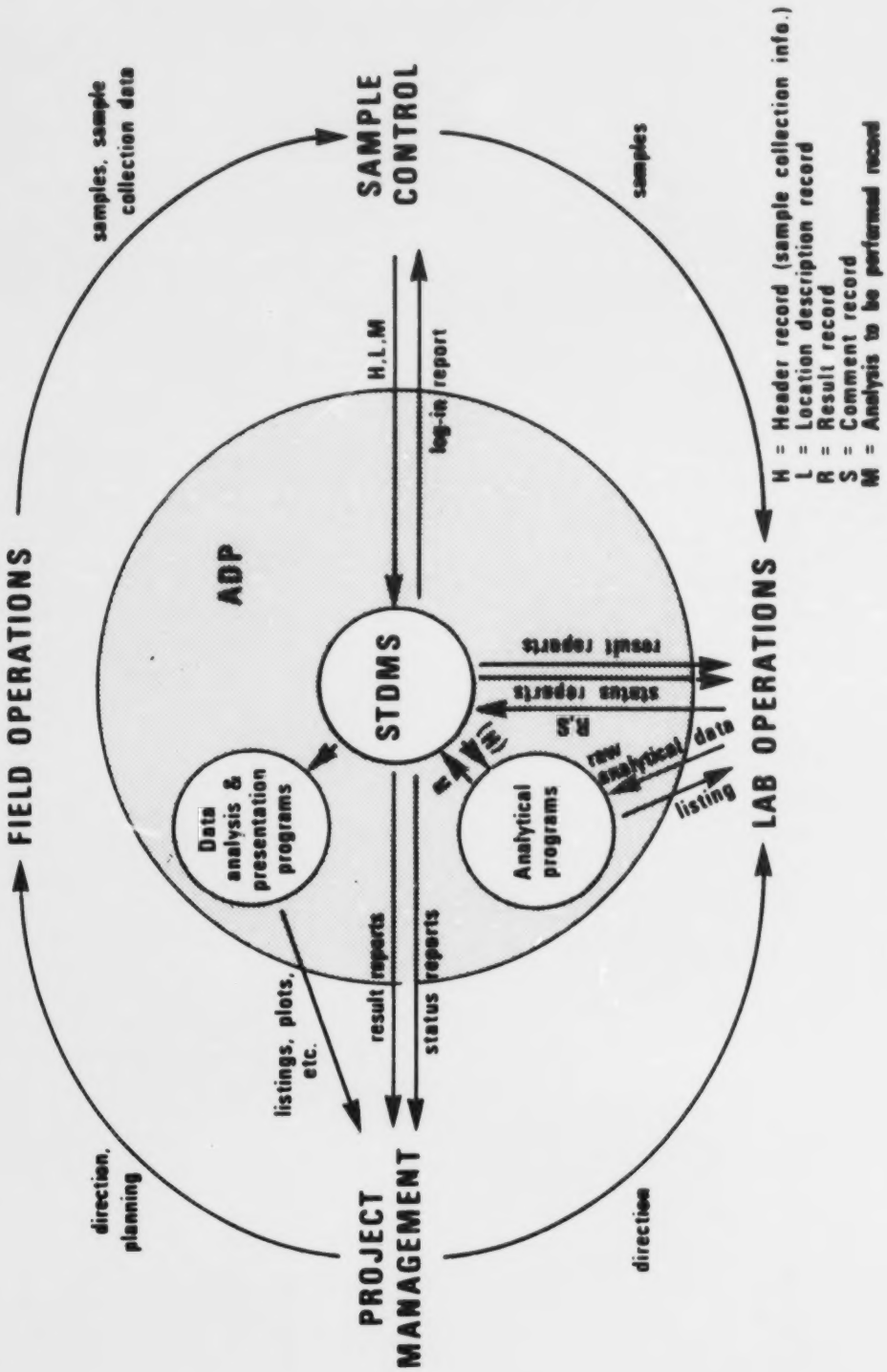
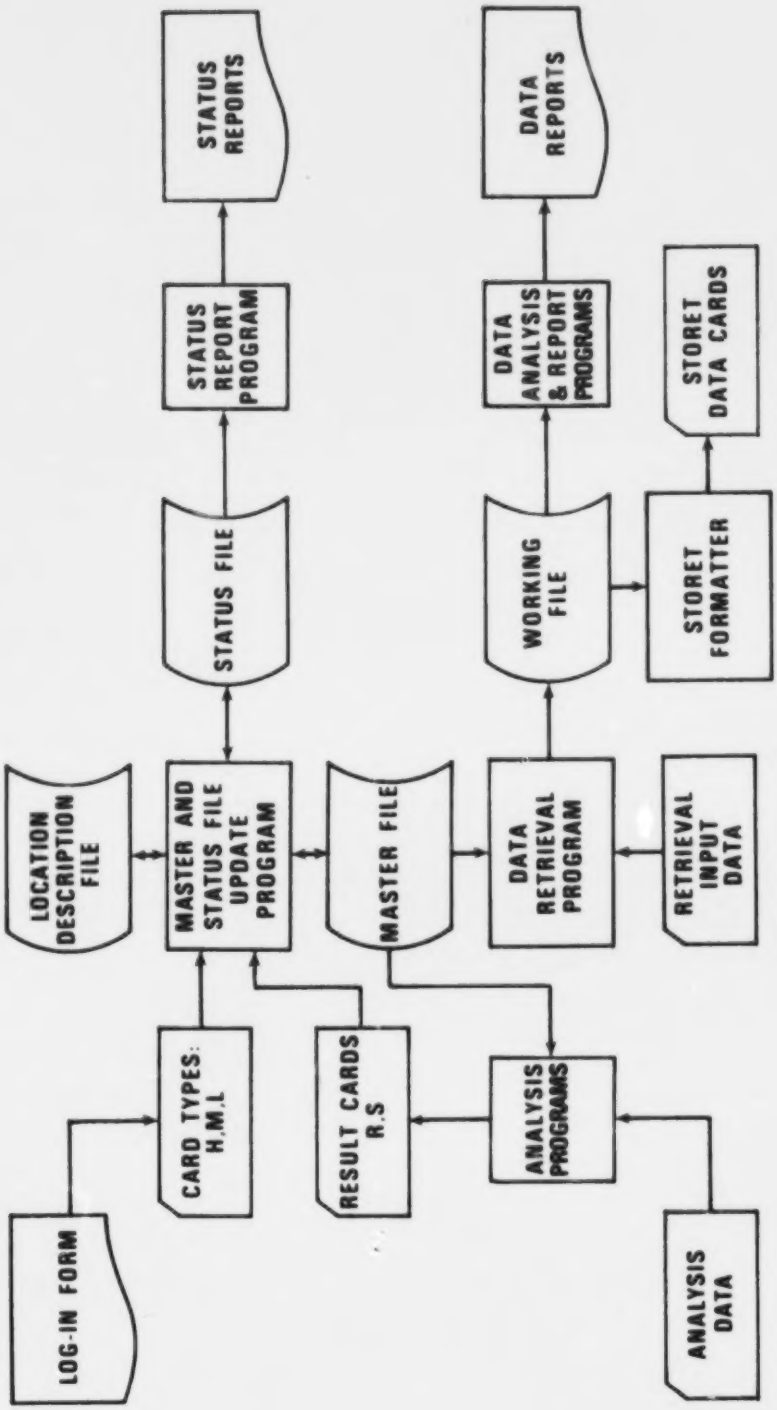
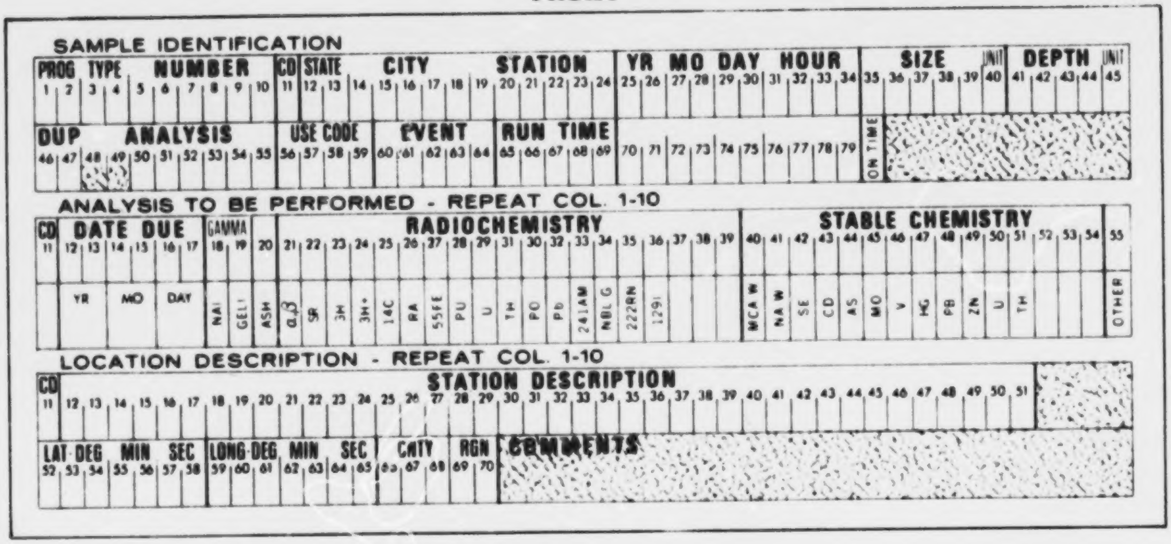


Figure 5
Sample Tracking Data Management System - Flow Diagram



FRONT



BACK

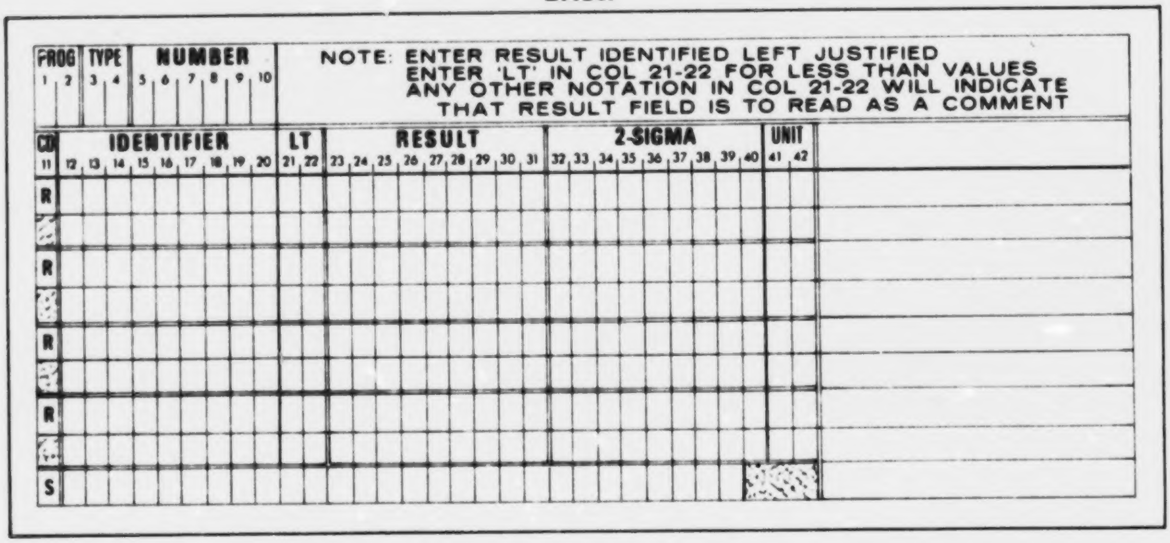


Figure 6
Sample Control Form

ANALYSIS IN PROCESS REPORT

PO ANALYSIS PAGE 1 REPORTING DATE 6/ 1/74

SAMPLE NO.	PROGRAM NO.	SAMPLE TYPE	LOGIN DATE	ANALYSIS DUE DATE	REPORT DUE DATE	NO. OF DAYS SINCE LOG IN
*129636	56	1	52074	50574	50574	12
*129650	56	1	52074	50774	50774	12
*129655	56	1	52074	50574	50574	12
*129657	56	1	52074	50574	50574	12
*129658	56	1	52074	51074	51074	12
*129761	56	1	52074	50974	50974	12
*129773	56	1	52074	51074	51074	12
*130030	20	1	52074	50674	50674	12
*130031	20	1	52074	50774	50774	12
*130035	20	1	52074	50574	50574	12
*130037	20	1	52074	50174	50174	12
*130038	20	1	52074	50474	50474	12
*130040	20	1	52074	50174	50174	12
*130041	20	1	52074	50374	50374	12
*130043	20	1	52074	51174	51174	12
*130044	20	1	52074	51174	51174	12
*130066	16	7	52074	51174	51174	12
*130072	16	7	52074	51474	51474	12
*130076	16	7	52074	51174	51174	12
*130078	16	7	52074	51374	51374	12
*130079	16	7	52074	51474	51474	12
*130082	16	7	52074	51174	51174	12
*130083	16	7	52074	51174	51174	12
*130084	16	7	52074	51274	51274	12
*130085	16	7	52074	51374	51374	12

* INDICATES ANALYSIS IS OVERDUE

Figure 7
Analysis in Process Report

ANALYSIS IN PROCESS
FOR
PROGRAM 51

CURRENT DATE 6/ 1/74 PAGE 2

SAMPLE NUMBER	ANALYSIS	SAMPLE TYPE	NO. OF DAYS SINCE LOG IN	ANALYSIS DUE DATE	REPORT DUE DATE
* 130557	GELI	3	12	5/ 3/74	5/ 3/74
* 130557	14C	3	12	5/ 3/74	5/ 3/74
* 130557	241AM	3	12	5/ 3/74	5/ 3/74
* 130557	MCA-W	3	12	5/ 3/74	5/ 3/74
* 130557	MG	3	12	5/ 3/74	5/ 3/74
* 130557		3	12	5/ 3/74	5/ 3/74
130558	PO	3	12	6/ 9/74	6/ 9/74
130558	MG	3	12	6/ 9/74	6/ 9/74
* 130559	14C	3	12	5/17/74	5/17/74
* 130559	241AM	3	12	5/17/74	5/17/74
* 130559		3	12	5/17/74	5/17/74
* 130560	NA-I	3	12	5/16/74	5/16/74
* 130560	GELI	3	12	5/16/74	5/16/74
* 130560	MCA-W	3	12	5/16/74	5/16/74
* 130560	MG	3	12	5/16/74	5/16/74
* 130561	14C	3	12	5/15/74	5/15/74
* 130561	PO	3	12	5/15/74	5/15/74
* 130561	241AM	3	12	5/15/74	5/15/74
* 130561	MG	3	12	5/15/74	5/15/74
* 130561		3	12	5/15/74	5/15/74
* 130562	GELI	3	12	5/10/74	5/10/74
* 130562	MG	3	12	5/10/74	5/10/74
* 130562		3	12	5/10/74	5/10/74
* 130563	GELI	3	12	5/11/74	5/11/74
* 130563	14C	3	12	5/11/74	5/11/74

* INDICATES ANALYSIS IS OVERDUE

Figure 8
Analysis in Process for Program 51

ANALYSIS STATUS REPORT
FOR
PROGRAM 51

SAMPLE NUMBER	SAMPLE TYPE	LOGIN DATE	NO. OF DAYS SINCE LOG IN	REPORT DUE DATE	NA.I	GEL I	ASH	A.H	SR	3H	3H+	14C	RA	5F	PU	UH	TH	PO	241AM	KR	XE	AR	RN	129I	210PB	MCA.W	NA.W	SE	CD	AS	MO	J	MG	PO	ZN	U	TH
130551	3	52074	12	5/23/74	I	.	I	I	.	I	O	.	O	O	.	O	O	.	O	O	.	O	O	.	O	O	.	O	O	.	O	O	.	O	I	.	I
130552	3	52074	12	6/10/74	I	.	I	I	.	I	I	.	I	I	.	I	I	.	C	I	.	I	I	.	I	C	.	I	I	.	I	I	.	I	I	.	I
130553	3	52074	12	5/11/74	O	.	O	O	.	O	O	.	O	O	.	O	O	.	O	O	.	O	I	.	I	O	.	I	I	.	I	I	.	I	O	.	O
130554	3	52074	12	6/17/74	I	.	I	I	.	I	I	.	I	I	.	I	I	.	C	I	.	I	I	.	I	C	.	I	C	.	C	C	.	C	I	.	I
130555	3	52074	12	5/18/74	O	.	I	I	.	I	I	.	I	I	.	I	O	.	O	O	.	O	I	.	I	O	.	I	I	.	I	I	.	I	I	.	I
130556	3	52074	12	5/ 1/74	O	.	O	O	.	O	O	.	O	O	.	O	I	.	I	I	.	I	I	.	I	O	.	I	I	.	I	I	.	I	O	.	O
130557	3	52074	12	5/ 3/74	I	.	I	I	.	I	I	.	I	I	.	I	I	.	O	I	.	I	I	.	I	O	.	I	I	.	I	I	.	I	I	.	I
130558	3	52074	12	6/ 9/74	C	.	C	C	.	C	C	.	C	C	.	C	I	.	I	I	.	I	C	.	C	C	.	C	I	.	I	I	.	I	C	.	C
130559	3	52074	12	5/17/74	O	.	O	O	.	O	I	.	I	I	.	I	I	.	O	I	.	I	O	.	O	O	.	O	O	.	O	O	.	O	I	.	I
130560	3	52074	12	5/16/74	O	.	I	I	.	I	O	.	O	O	.	O	O	.	O	O	.	O	I	.	I	O	.	I	I	.	I	I	.	I	O	.	O
130561	3	52074	12	5/15/74	O	.	O	O	.	O	I	.	I	I	.	I	I	.	O	I	.	I	O	.	O	O	.	O	I	.	I	I	.	I	I	.	I
130562	3	52074	12	5/10/74	I	.	I	I	.	I	O	.	O	O	.	O	O	.	O	O	.	O	O	.	O	O	.	O	I	.	I	I	.	I	I	.	I
130563	3	52074	12	5/11/74	I	.	I	I	.	I	I	.	I	I	.	I	I	.	O	I	.	I	I	.	I	O	.	I	O	.	O	O	.	O	I	.	I
130564	3	52074	12	5/12/74	O	.	I	I	.	I	I	.	I	I	.	I	I	.	O	I	.	I	I	.	I	O	.	I	I	.	I	I	.	I	I	.	I

I - INCOMPLETE ANALYSIS

C - COMPLETED ANALYSIS

O - OVERDUE ANALYSIS

. - NOT REQUESTED

Figure 9
Analysis Status Report for Program 51

SIDES

By David Barrow

The STORET Input Data Editing System (SIDES) was designed and built by several members of the Region IV staff of the Southeast Environmental Research Laboratory to overcome some difficulties in using STORET. These difficulties are outlined in Figure 1. Figure 2 gives a list of features that SIDES provides to alleviate these difficulties.

Throughout the past two years, SIDES has proven to be a useful vehicle for preparing data for entry into STORET and for obtaining data printouts for technical publications much sooner than they could have been obtained after entry of the data into STORET.

Figure 3 gives a list of users and uses of the SIDES system. In addition to its primary use for data entry, the edit program of SIDES is being used as a data editing/interface module in the transfer of Ohio River Valley Sanitation Commission's (ORSANCO) water users' data from ORSANCO's data files into STORET. The SIDES edit program is also being used in the Office of Research and Development's Interim Laboratory Data Management System as the edit module and as an interface to STORET.

Figure 4 shows the system block diagram for the SIDES system. It also shows a program Loadings and Statistics (LDSTAT) and its interface to the SIDES data files. While program LDSTAT is not officially a part of the SIDES system, it provides a sophisticated printout capability for data prior to entry into STORET. Program LDSTAT also may be used to print out data already stored in STORET.

The SIDES system block diagram begins by showing two card decks: the editing information deck (EID) and the data decks. Figure 5 gives an example of these two decks. The EID consists of a STORET agency card, a YEAR card, a parameter (PAR) card for each parameter that may have data, and a STATN card for each station that may have data. The function of the EID is simply to provide an easy reference for the edit program. The edit program checks much of the contents of the data decks against what was specified in the EID.

The second card deck, the data decks, consists of one or more data decks. Each data deck in turn consists of one parameter (PAR) card and any number of sample identification cards, type B (SIDB) followed by corresponding data cards (DTA). The PAR card specifies the order and parameter number of the data punched into the DTA cards that follow it.

The SIDB card contains the sample identification information for each sample to be processed by SIDES. The following information may be specified on the SIDB card: STORET station number; the date and time the sample was taken; the ending date-time and the various composite description information for composite samples; the depth at which the sample was taken; and the percent from the right bank of the location at which the sample was taken.

The DTA card gives the actual values measured for each parameter specified on the PAR card. The DTA card is linked to its SIDB card by the sample number which must appear on both. The DTA card may be continued on up to nine additional cards. At six parameter values/card, this gives a total of 60 parameters that may be processed through the SIDES system at one time.

If the user has data for more than 60 parameters, he must separate the data into two or more groups of 60 or less parameters each and process the groups in separate runs of the system. On the DTA card, a missing parametric value is indicated by a blank field, and the deletion of a value already in STORET is indicated by a solitary D in the first position of the proper field.

The EID and Data Decks are processed through the merge/edit (M/E) phase of the system during which several functions are being performed. First the SIDB and DTA cards are merged. Second, extensive editing is performed upon the cards of the data decks. All cards in both the EID and the data decks are printed on the M/E printout. If any errors are detected during the editing operation, appropriate error messages are printed out with the guilty card on the M/E printout. Examples of M/E printouts appear in Figures 6, 7, and 8.

In the third step of the operation, the data in the data decks are reformatted and entered into the SIDES disk files. All samples are entered into the file that goes to the SIDES printout programs, but all samples having errors are flagged. Only samples that passed all edit tests

are entered into the file that goes to the STORET storage program.

Once the M/E phase has been completed, the user has many options. If errors were detected during the M/E phase, he may decide to correct the errors and rerun the M/E phase. He may decide to store the good data, correct the erroneous data and rerun it. He may decide to obtain an original order or a station order printout of the data to further assist him in validating the data; or he may decide to get a LDSTAT printout to obtain statistics, loadings, or a printout to go into a technical report.

An example of the original order printout appears in Figures 9 and 10. An example of the station order printout appears in Figures 11 and 12. In the original order and station order printouts, the parameter values are printed exactly as they were punched onto the DTA cards. This is to facilitate checking the data. Note that even the samples that were rejected in the M/E printouts are included. The original order printout greatly facilitates proofreading the data from the original input documents. The station order printout enables chemistry personnel to easily discover bad values that may have occurred during transcription.

An example of the LDSTAT printout is given in Figures 13 and 14. In the LDSTAT printout, only samples that have passed all edit tests are printed. The parameter values are printed with decimal points aligned as dictated by the STORET decimal point specification for each parameter. The statistics are evaluated for one station at a time. Note that the remarks K and L are included in the minimum and maximum values. Statistics are not printed for parameters having only one value. Some parameters are not summarized at all, e.g., SAMPLE NUMBER.

Why SIDES was written:

- STORET input formats (Fixed Format and DIP) were difficult to code and to keypunch thereby causing errors and wasted time.
- There were no easy-to-read printouts to facilitate validation of data prior to entry into STORET.
- The length of time required to enter data and corrections into STORET was unsatisfactory.
- It was very difficult to correct data once it had been entered into STORET.

Figure 1
SIDES

SIDES offers the following features:

- An easy-to-use card format for entering data particularly from the standpoint of keypunching.
- A straightforward method of entering sample identification information.
- A comprehensive pre-STORET editing system to allow user to catch keypunch and transcription errors prior to entering the data into STORET.
- Several data reports may be run from SIDES intermediate disk files to aid in data validation and to provide data printouts with statistics for inclusion in technical reports prior to entry into STORET.
- A simplified method of correcting data already entered into STORET through SIDES.

Figure 2
Features of SIDES

SIDES users and uses:

- State of North Carolina for data entry
- State of South Carolina for data entry
- EPA, Region IV, S&A division, for:
  - Data entry
  - Technical report generation
  - ORSANCO water users' data transfer
- EPA, Region IV, Escambia Bay Recovery Study for data entry
- EPA, ORD, Quality Assurance Division, for:
  - Part of the Interim Laboratory Data Management System.

Figure 3
SIDES Uses

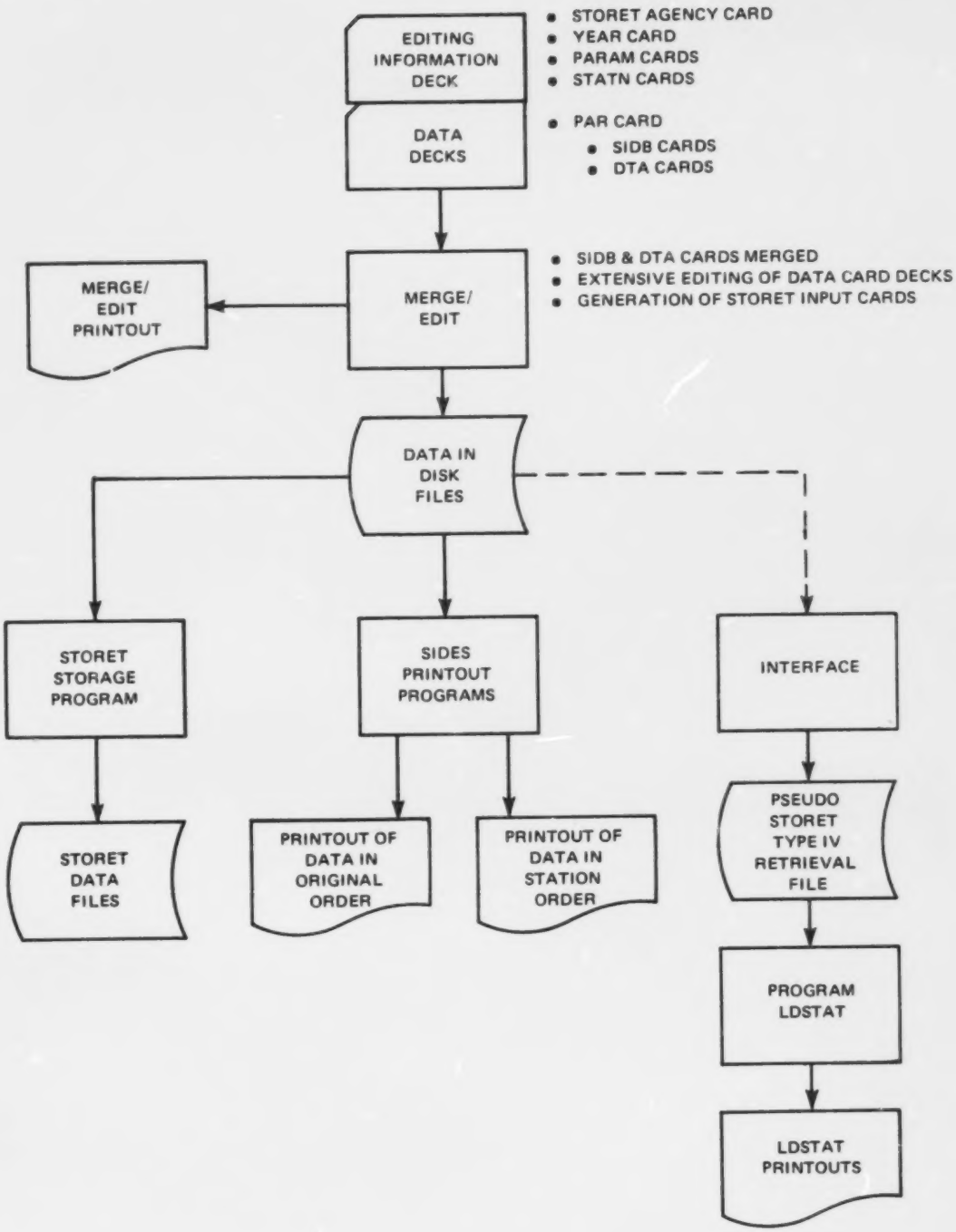


Figure 4
SIDES System Block Diagram

```
**********        EDITING INFORMATION DECK FOLLOWS.        **********

11135000                  SEWL ATHENS GA 404-546-3548          F   021112049999  A
YEAR      74
STATN     WC-1       130000
STATN     S-1        130005
PARAM     31501
PARAM     31616
PARAM     00400
PARAM     00300
PARAM     00940
PARAM     00900
PARAM     00410
PARAM     50060
PARAM     00340
PARAM     00680



**********        ONE DATA CARD DECK FOLLOWS.        **********
          |    |    |    |    |    |    |    |    |
PAR    31501316160040000300009400090000410500600034000680
STDB      WC-1                 740521 0900       740521 1100          0001
STDB      S-1                  740515 1000                            0005
STDB      WC-1                 740521 0905                            0002
STDB      S-1                  740515 1006                            0011
STDB      WC-1                 740521 0945                            0003
STDB      S-1                  740515 1011                            0015
STDB      WC-1                 740521 1030                            0004
STDB      S-2                  740515 1001                            0006
STDB      S-1                  740229 1002                            0007
STDB      S-1                  740515 1063                            0008
STDB      S-1                  740515 1004                            0009
STDB      S-1                  740515 1005                            0010
STDB      S-1                  740515 1007                            0012
STDB      S-1                  740515 1008                            0013
STDB      S-1                  740515 1009                            0014
STDB      S-1                  740515 1012                            0016
STDB      S-1                  740515 1013                            0017
STDB      S-1                  740515 1014                            0018
DTA     0001     17000      740
DTA   1 0005                          144        100
DTA     0002                          6.2        8.5        1.9        13
DTA   1 0002     6          K0.1
DTA   1 0011                          154
DTA     0003                          6.3        8.3        2.0        15
DTA   1 0003     7          K0.1
DTA   1 0015                          154
DTA     0004                          6.3        8.1        2.0        15
DTA   1 0004     7          K0.1
DTA   1 0006                          136
DTA   1 0007                          148
DTA   1 0008                          140
DTA   1 0009               29.3                  150F6
DTA   1 0010     15         2 5        141         25
DTA   1 0012                          148
DTA   1 0013                          154
DTA   1 0014                          152
DTA   1 0016                          136
DTA   1 0017                          152
DTA   1 0018                          169                                (S.2.0)
```

Figure 5
Example of Two Card Deck

```
REPORT RFST01A                                                                                                                  PAGE    1

EDITING INFORMATION FILE:

   10 '1113S000                  SEWL ATHENS GA 404-546-3548          F   021112049999  A'
   20 'YEAR      74                                                                        '
   30 'STATN     WC-1       130006                                                         '
   40 'STATN     S-1        130005                                                         '
   50 'PARAM     31501                                                                     '
   60 'PARAM     31616                                                                     '
   70 'PARAM     00400                                                                     '
   80 'PARAM     00300                                                                     '
   90 'PARAM     00940                                                                     '
  100 'PARAM     00900                                                                     '
  110 'PARAM     00410                                                                     '
  120 'PARAM     50060                                                                     '
  130 'PARAM     00340                                                                     '
  140 'PARAM     00680                                                                     '
```

Figure 6
Merge/Edit Printout
Field Study Data Edit Report

```
REPORT RFST01A                                                                                                      PAGE    2

FIELD STUDY DATA DECK FOLLOWS:

  10 *PAR     3150131616004000030000940009000041050060003400068O                     *
  20 *SIDB     WC-1              740521 0900      740521 1100        0001            *
  30 *DTA     0001    17000      790                                                 *                               COMP-TIME
  40 *SIDB     S-1               740515 1000                         0005            *
  50 *DTA   1 0005                      144      100                                 *                               GRAB
  60 *SIDB     WC-1              740521 0905                         0002            *
  70 *DTA     0002                      6.2      8.5       1.9       13              *
  80 *DTA   1 0002    6         K0.1                                                 *                               GRAB
  90 *SIDB     S-1               740515 1006                         0011            *
 100 *DTA   1 0011                      154                                          *                               GRAB
 110 *SIDB     WC-1              740521 0945                         0003            *
 120 *DTA     0003                      6.3      8.3       2.0       15              *
 130 *DTA   1 0003    7         K0.1                                                 *                               GRAB
 140 *SIDB     S-1               740515 1011                         0015            *
 150 *DTA   1 0015                      154                                          *                               GRAB
 160 *SIDB     WC-1              740521 1030                         0004            *
 170 *DTA     0004                      6.3      8.1       2.0       15              *
 180 *DTA   1 0004    7         K0.1                                                 *                               GRAB
 190 *SIDB     S-2               740515 1001                         0006            *-->ILL.FIELD STA.NR.
 200 *DTA   1 0006                      136                                          *-->BAD SIDB CARD.              REJECTED
 210 *SIDB     S-1               740229 1002                         0007            *-->ST.DA.OUT OF RANGE.
 220 *DTA   1 0007                      148                                          *-->BAD SIDB CARD.              REJECTED
 230 *SIDB     S-1               740515 1063                         0008            *-->ST.TIME OUT OF RANGE.
 240 *DTA   1 0008                      140                                          *-->BAD SIDB CARD.              REJECTED
 250 *SIDB     S-1               740515 1004                         0009            *
 260 *DTA   1 0009            29.3               150E6                               *-->FLD  2 ILL.CHAR.BET.FLDS.   REJECTED
 270 *SIDB     S-1               740515 1005                         0010            *
 280 *DTA   1 0010    15          2 5   141       25                                 *-->FLD  1 ILL.VALUE.
                                                                                      -->FLD  2 ILL.BLANK.
                                                                                      -->FLD  4 NOT LEFT JUSTIFIED.  REJECTED
 290 *SIDB     S-1               740515 1007                         0012            *
 300 *DTA   1 0012                      148                                          *                               GRAB
 310 *SIDB     S-1               740515 1008                         0013            *
 320 *DTA   1 0013                      154                                          *                               GRAB
 330 *SIDB     S-1               740515 1009                         0014            *
 340 *DTA   1 0014                      152                                          *                               GRAB
 350 *SIDB     S-1               740515 1012                         0016            *
 360 *DTA   1 0016                      136                                          *                               GRAB
 370 *SIDB     S-1               740515 1013                         0017            *
 380 *DTA   1 0017                      152                                          *                               GRAB
 390 *SIDB     S-1               740515 1014                         0018            *
 400 *DTA   1 0018                      169                                          *                               GRAB
```

Figure 7
Merge/Edit Printout
Field Study Data Edit Report

```
REPORT RFST01A                                                          PAGE    3

EOF ON CFST01.


PROGRAM PFST01 RUN SUMMARY:

            40 CARDS READ FROM CFST01.
            30 CARDS WRITTEN ONTO DFST03 FOR STORAGE IN STORET.
            44 RECORDS WRITTEN ONTO DFST04.

          STORET INPUT HAS BEEN GENERATED FOR
                 12 GRAB SAMPLES.
                  0 COMPOSITE SAMPLES W/O TIME OF DAY.
                  1 COMPOSITE SAMPLES WITH TIME OF DAY.
                  0 SAMPLE DELETIONS.

          AND DATA FOR    5 SAMPLES HAS BEEN REJECTED.


EXECUTION COMPLETED.
```

Figure 8
Merge/Edit Printout
Field Study Data Edit Report

PAGE 1

FIELD STA.NR.	STORET STA.NR.	FIELD SAMPLE IDENTIFICATION START DATE	ST. TIME	COMP.. TYPE	ENDING DATE	END. TIME	SAMPLE NUMBER	00008 SEWL SAMPLE NUMBER	00003 DEPTH	00002 % FROM RIGHT BANK	31501 TOT COLI MFIMENDO /100ML	31616 FEC COLI MFM-FCBR /100ML	00400 PH SU	00300 DO MG/L
WC-1	130000	740521	0900		740521	1100	0001	0001			17000	790		
S-1	130005	740515	1000				0005	0005					6.2	8.5
WC-1	130000	740521	0905				0002	0002						
S-1	130005	740515	1006				0011	0011					6.3	8.3
WC-1	130000	740521	0945				0003	0003						
S-1	130005	740515	1011				0015	0015					6.3	8.1
WC-1	130000	740521	1030				0004	0004						
S-2		740515	1001				0006	0006						
S-1	130005	740229	1002				0007	0007						
S-1	130005	740515	1063				0008	0008						
S-1	130005	740515	1004				0009	0009						
S-1	130005	740515	1005				0010	0010						
S-1	130005	740515	1007				0012	0012						
S-1	130005	740515	1008				0013	0013						
S-1	130005	740515	1009				0014	0014						
S-1	130005	740515	1012				0016	0016						
S-1	130005	740515	1013				0017	0017						
S-1	130005	740515	1014				0018	0018						

Figure 9
Original Order
Parameters 1 Through 7

FIELD STA.NR.	STORET STA.NR.	FIELD SAMPLE IDENTIFICATION START DATE	ST. TIME	COMP.. TYPE	ENDING DATE	END. TIME	SAMPLE NUMBER	00940 CHLORIDE CL MG/L	00900 TOT HARD CACO3 MG/L	00410 T ALK CACO3 MG/L	50060 CHLORINE TOT RESD MG/L	00340 COD HI LEVEL MG/L	00680 T ORG C C MG/L
WC-1	130000	740521	0900		740521	1100	0001						
S-1	130005	740515	1000				0005					144	100
WC-1	130000	740521	0905				0002	1.9	13	6	K0.1		
S-1	130005	740515	1006				0011					154	
WC-1	130000	740521	0945				0003	2.0	15	7	K0.1		
S-1	130005	740515	1011				0015					154	
WC-1	130000	740521	1030				0004	2.0	15	7	K0.1		
S-2		740515	1001				0006						
S-1	130005	740229	1002				0007						
S-1	130005	740515	1063				0008						
S-1	130005	740515	1004				0009				9.3		150E6
S-1	130005	740515	1005				0010			15	2 5	141	
S-1	130005	740515	1007				0012					148	
S-1	130005	740515	1008				0013					154	
S-1	130005	740515	1009				0014					152	
S-1	130005	740515	1012				0016					136	
S-1	130005	740515	1013				0017					152	
S-1	130005	740515	1014				0018					169	

Figure 10
Original Order
Parameters 8 Through 14

FIELD STA.NR.	STORET STA.NR.	FIELD SAMPLE IDENTIFICATION START DATE	ST. TIME	COMP.. TYPE	ENDING DATE	END. TIME	SAMPLE NUMBER	00008 SEWL SAMPLE NUMBER	00003 DEPTH	00002 % FROM RIGHT BANK	31501 TOT COLI MFIMENDO /100ML	31616 FEC COLI MFM-FCBR /100ML	00400 PH SU	00300 DO MG/L
S-2		740515	1001				0006	0006						
WC-1	130000	740521	0945				0003	0003						
WC-1	130000	740521	0905				0002	0002					6.3	8.3
WC-1	130000	740521	0900		740521	1100	0001	0001					6.2	8.5
WC-1	130000	740521	1030				0004	0004			17000	790		
													6.3	8.1
S-1	130005	740229	1002				0007	0007						
S-1	130005	740515	1011				0015	0015						
S-1	130005	740515	1000				0005	0005						
S-1	130005	740515	1006				0011	0011						
S-1	130005	740515	1014				0018	0018						
S-1	130005	740515	1013				0017	0017						
S-1	130005	740515	1012				0016	0016						
S-1	130005	740515	1009				0014	0014						
S-1	130005	740515	1008				0013	0013						
S-1	130005	740515	1007				0012	0012						
S-1	130005	740515	1005				0010	0010						
S-1	130005	740515	1004				0009	0009						
S-1	130005	740515	1063				0008	0008						

Figure 11
Station Order
Parameters 1 Through 7

FIELD STA.NR.	STORET STA.NR.	FIELD SAMPLE IDENTIFICATION START DATE	ST. TIME	COMP.. TYPE	ENDING DATE	END. TIME	SAMPLE NUMBER	00940 CHLORIDE CL MG/L	00900 TOT HARD CACO3 MG/L	00410 T ALK CACO3 MG/L	50060 CHLORINE TOT RESD MG/L	00340 COD HI LEVEL MG/L	00680 T ORG C C MG/L
S-2		740515	1001				0006						
WC-1	130000	740521	0945				0003	2.0	15	7	K0.1		
WC-1	130000	740521	0905				0002	1.9	13	6	K0.1		
WC-1	130000	740521	0900		740521	1100	0001						
WC-1	130000	740521	1030				0004	2.0	15	7	K0.1		
S-1	130005	740229	1002				0007					154	
S-1	130005	740515	1011				0015					144	100
S-1	130005	740515	1000				0005					154	
S-1	130005	740515	1006				0011				9.3		150E6
S-1	130005	740515	1004				0009						
S-1	130005	740515	1063				0008					148	
S-1	130005	740515	1007				0012			15	2 5	141	
S-1	130005	740515	1005				0010					152	
S-1	130005	740515	1009				0014					154	
S-1	130005	740515	1008				0013					152	
S-1	130005	740515	1013				0017					136	
S-1	130005	740515	1012				0016					169	
S-1	130005	740515	1014				0018						

Figure 12
Station Order
Parameters 8 Through 14

STATION - 130000

DATE	TIME	DATE	TIME	00008 SEWL SAMPLE NUMBER	00940 CHLORIDE CL MG/L	00900 TOT HARD CACO3 MG/L	00410 T ALK CACO3 MG/L	50060 CHLORINE TOT RESD MG/L	31501 TOT COLI MFIMENDO /100ML	31616 FEC COLI MFM-FCBR /100ML	00400 PH SU	00300 DO MG/L
740521	0900 (C)	740521	1100	1					17000	790		
		740521	0905	2	1.9	13	6	0.10K			6.2	8.5
		740521	0945	3	2.0	15	7	0.10K			6.3	8.3
		740521	1030	4	2.0	15	7	0.10K			6.3	8.1
740521												
NUMBER					3	3	3	3	1	1	3	3
MAXIMUM					2.0	15	7	0.10K			6.3	8.5
MINIMUM					1.9	13	6	0.10K			6.2	8.1
SUM					5.9	43	20	0.30			18.8	24.9
SUM SQ.					11.6	619	134	0.03			117.8	206.7
MEAN					2.0	14	7	0.10			6.3	8.3
VARIANCE					0.0	1	0	0.00			0.0	0.0
STD.DEV.					0.1	1	1	0.00			0.1	0.2
STD.ERR.					0.0	1	0	0.00			0.0	0.1
COEF VAR					2.9	8	9	0.00			0.9	2.4
LOG MEAN					2.0	14	7	0.10			6.3	8.3
740521												

Figure 13
LDSTAT Printout
Test Printout of SIDES Data

STATION - 130005

DATE	TIME	DATE	TIME	00008 SEWL SAMPLE NUMBER	00340 COD HI LEVEL MG/L	00680 T ORG C C MG/L
		740515	1000	5	144	100.0
		740515	1006	11	154	
		740515	1007	12	148	
		740515	1008	13	154	
		740515	1009	14	152	
		740515	1011	15	154	
		740515	1012	16	136	
		740515	1013	17	152	
740515						
NUMBER					8	1
MAXIMUM					154	
MINIMUM					136	
SUM					1194	
SUM SQ.					178492	
MEAN					149	
VARIANCE					41	
STD.DEV.					6	
STD.ERR.					2	
COEF VAR					4	
LOG MEAN					149	
740515						

Figure 14
LDSTAT Printout
Test Printout of SIDES Data

SELECTION OF PROBABILITY MODELS FOR DETERMINING QUALITY CONTROL DATA SCREENING RANGE LIMITS

By Wayne R. Ott, Ph.D.

INTRODUCTION

Environmental monitoring laboratories performing routine chemical analyses are using computers increasingly to process and store the data. If these laboratories process large quantities of data, errors may be introduced in the data handling phase, and there is some risk that these errors may go undetected. Such errors, which frequently result simply from keypunch mistakes, may be stored as "valid" data and may create serious problems when sample means or other statistics are calculated. The computer can be used, however, to automatically screen laboratory data as they are entered to determine if any value of a given parameter* is "unusual" or lies outside an "acceptable range." Such unusual values can be "flagged" by the computer, thereby bringing them to the attention of laboratory personnel where errors can be corrected and new analyses performed if necessary.

STATEMENT OF PROBLEM

Typical environmental data processing and handling errors result from three sources: keypunch mistakes, failure to consider minimum detectable limits, and clerical mistakes or omissions. For example, when storing water quality data, a valid parameter value of pH = 5.5 may be erroneously keypunched as pH = 55.0 simply because the decimal is entered in the wrong column. Without provision for "data screening," the computer will simply accept this value and store it as valid data. To screen the data, the value of each parameter is compared with preset limits which are also stored in the computer; values outside these limits are flagged. Values which are improbable can therefore be brought to the chemist's attention while impossible values can be automatically rejected.

To implement a data screening system, an improbable range of values for each parameter must be determined; historical data can be used for this purpose. At first glance, two or three times the standard deviation† might appear to be a suitable choice for the quality control data screening "range limits." However, an examination of water quality parameters suggested that most parameters are not normally distributed, and the underlying distribution of each parameter may differ. Thus, use of 2 σ or 3 σ would result in range limits that have different probability of occurrence for each parameter. It would be preferable for the range limits to always represent the same probability, regardless of the underlying distribution. Furthermore, it would be desirable to have a systematic procedure for arriving at these data screening range limits.

METHODOLOGY

The Environmental Protection Agency (EPA) is now undertaking a data analysis effort in which probability models are fit to historical water quality data and range limits calculated. These limits then are being built into a quality control data screening computer program which serves as an integral part of a computerized laboratory data management system. This data management system is designed to store, manipulate, and retrieve environmental data, ultimately transferring it to national environmental data archives.

---

*In this paper, the term "parameter" is used to mean a variable which is used as a measure of the state of the environment. For example, both carbon monoxide measured by air monitoring stations and pH measured in rivers are environmental parameters.

†The data estimate $\bar{x}$ of the arithmetic mean μ_1' is calculated as follows: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

The data estimate s of the standard deviation σ is calculated as follows: $s = \left\{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2\right\}^{1/2}$

Probability Models

Most environmental parameters, such as pollutant concentration, never are less than zero and have essentially unbounded upper limits. For environmental data, eight probability models are of particular interest:*

- Exponential
- Normal
- Log-normal
- Weibull
- Gamma
- Beta
- Raleigh
- Extreme Value.

Equations (1) and (2) give the probability density function (PDF) and the cumulative distribution function (CDF) for the Weibull probability model:

PDF

$$f_X(x) = \frac{\beta}{\xi}\left(\frac{x}{\xi}\right)^{\beta-1} e^{-\left[\frac{x}{\xi}\right]^{\beta}}, \quad x \geq 0 \qquad (1)$$

CDF

$$F_X(x) = 1 - e^{-\left[\frac{x}{\xi}\right]^{\beta}} \qquad (2)$$

This probability model has two parameters: a "shape parameter," β, and a "scale parameter," ξ. For different values of β, it takes on a variety of different shapes (Figure 1). Because of this flexibility, the model is a good candidate for fitting environmental data. When $\beta = 1$, the Weibull probability model becomes equivalent to the exponential probability model,

$$f_X(x) = \frac{1}{\xi} e^{-x/\xi} \qquad (3)$$

The other probability models differ in shape from each other. Some, like the normal distribution, are symmetrical. The characteristic shape is constant, and the parameters determine the "location" of the curve and its "spread" or dispersion.

Overall Approach

A practical, convenient approach for fitting probability models to environmental data is by the "method of moments." In this approach, the estimates of the moments calculated from the data are set equal to the actual moments of the probability model. The overall methodology for selecting and fitting models to historical data uses this approach and is depicted in Figure 2. At the left side of the figure, the raw historical data are analyzed, by computer, and the histogram and estimates of the moments are developed. Next, a probability model is selected for trial. The relative magnitude of the moments gives some clue as to the most appropriate model. The cumulative distribution function of the model is then compared with the cumulative frequencies of the data. If the model does not have a reasonably good fit, another model is selected. Once a "best-fit" model is selected, it is used to calculate, within the range of the data, the values of the parameter which correspond to various extreme probabilities (P = .01, .001; P = .99, .999). These values are used as the data screening range limits.

Selection of Probability Models

The most important distinction between different probability models is their shape. The third and fourth moments provide measures of the shape of each model. The third moment, μ_3, is a measure of the distribution's asymmetry or *skewness*, the extent to which it exhibits a long tail to the left or to the right. The fourth moment is a measure of the distribution's "peakedness" or *kurtosis*, the relationship of the distribution's "height" to its "width." For a given probability model, the second, third, and fourth moments are obtained from the probability density function $f_X(x)$ as follows, where μ_1' is the mean:

$$\mu_2 = \int_{-\infty}^{\infty} \left(x - \mu_1'\right)^2 f_X(x)\, dx \qquad (4)$$

$$\mu_3 = \int_{-\infty}^{\infty} \left(x - \mu_1'\right)^3 f_X(x)\, dx \qquad (5)$$

$$\mu_4 = \int_{-\infty}^{\infty} \left(x - \mu_1'\right)^4 f_X(x)\, dx \qquad (6)$$

The second moment is also known as the variance and is the square of the standard deviation, $\mu_2 = \sigma^2$. Two

*Gerald J. Hahn and Samuel S. Shapiro, *Statistical Models in Engineering* (New York: John Wiley & Sons, Inc., 1967). See pages 120 through 134 for summary of probability models for continuous random variables.

important quantities are the coefficient of skewness $\sqrt{\beta_1}$ and the coefficient of kurtosis β_2:

$$\sqrt{\beta_1} = \frac{\mu_3}{(\mu_2)^{3/2}} \tag{7}$$

$$\beta_2 = \frac{\mu_4}{(\mu_2)^2} \tag{8}$$

Figure 3 shows the relative values of β_1 and β_2 calculated by the author for a variety of probability models.* The uniform, normal, Raleigh, and exponential models all have constant shapes; therefore, they are represented on this figure as points. For the normal distribution, $\beta_1 = 0$ (indicating that the model is symmetrical) and $\beta_2 = 3$. The Gamma model is represented as a straight line, since a linear relationship exists between β_1 and β_2:

$$\beta_2 = 1.5\,\beta_1 + 3 \tag{9}$$

On this figure, both the lines for the Gamma and Weibull probability models intersect the point representing the exponential probability model because the exponential model is one special case of both of these distributions. For the log-normal and Weibull probability models, the lines, which were calculated by computer, exhibit slight curvature. The beta distribution (not shown) would be represented as a region on this figure rather than a line since it takes on a multitude of shapes.

For a given set of data, whose underlying distribution may be unknown, it is possible to obtain estimates of the first four moments. Estimates of the first and second moments are calculated as the mean $\bar{x}$ and variance s^2 in the manner given previously. Estimates of the third and fourth moments about the mean are obtained as follows:

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^3 \tag{10}$$

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^4 \tag{11}$$

Using these quantities, estimates of the coefficients of skewness and kurtosis may be obtained by substituting s^2 for m_2, m_3 for μ_3, and m_4 for μ_4 into equations (7) and (8):

$$\sqrt{b_1} = \frac{m_3}{s^3} \tag{12}$$

$$b_2 = \frac{m_4}{s^4} \tag{13}$$

These quantities, which are easily calculated by computer from the raw data, can assist in selecting the most appropriate probability model. Once calculated, they may be compared to the curves and points in the $(\beta_1\ \beta_2)$ plane of Figure 3 to suggest which probability model is a good choice. By establishing a "region of acceptance" around each line and point in the $(\beta_1\beta_2)$ plane, this comparison can be done automatically by computer.

APPLICATION OF METHODOLOGY

For the laboratory data management system, three classes of suspect data were selected:

Improbable:
$P(X \leq x_L \text{ or } X \geq x_H) = .01$

Highly improbable:
$P(X \leq x_L \text{ or } X \geq x_H) = .001$

Impossible:
Cannot physically occur.

The probability model is used to calculate x_L and x_H for the first two classes. With the exception of pH, most of the water quality parameters had *impossible* lower ranges of 0.0 and no upper impossible range. Another important criterion for impossible values was based on interparameter comparisons, the physical and chemical relationships among one or more environmental parameters; but these factors will not be discussed here.

To illustrate the method, atmospheric concentrations of lead in cities throughout the United States in 1968 will be used as an example. The computer-generated histogram of these 4411 values is shown in Figure 4. This histogram is greatly skewed to the right, with an arithmetic mean of 0.862 micrograms per cubic meter ($\mu g/m^3$). For these data $b_1 = 13.9$ and $b_2 = 25.8$. In this case, the skewness of the raw data and the exponential model differs, as does the kurtosis, but the

*This figure is patterned after the (β_1, β_2) plane by E.S. Pearson, University College, London, cited by Hahn and Shapiro, *op. cit.*, p.197. All curves have been recalculated.

exponential model offers advantages due to its simplicity, and we shall apply it in this instance for purposes of illustration.

The exponential probability model has just one parameter, ξ, as seen in equation (3), which is also its mean. We use the method of moments to set $\xi = 0.862$ $\mu g/m^3$, the mean calculated from the data. The cumulative distribution function $F_X(x)$ for this model then becomes:

$$F_X(x) = 1 - e^{-x/.862} \qquad (14)$$

The plot of this probability model and the histogram of the data are both shown in Figure 5. Using equation (13), the range limits for atmospheric lead concentrations are then calculated as the values for which $F_X(x) = 0.001, 0.01, 0.99$, and 0.999. A sample calculation for $F_X(x) = 0.99$ is as follows:

$$F_X(x) = 1 - \exp[-x/.862] = 0.99$$

$$\exp[-x/.862] = 0.01$$

$$-x/.862 = \ln(0.01)$$

$$x = (-0.862)(-4.605) = 3.97\ \mu g/m^3$$

The resulting range limits are obtained in this same manner:

Improbable

$$F_X(x_L) = 0.01 \qquad x_L = .0087\ \mu g/m^3$$

$$F_X(x_H) = 0.99 \qquad x_H = 3.97\ \mu g/m^3$$

Highly Improbable

$$F_X(x_L) = 0.001 \qquad x_L = .00086\ \mu g/m^3$$

$$F_X(x_H) = 0.999 \qquad x_H = 5.95\ \mu g/m^3$$

These limits, which are based on national data, then can be programmed into the computer as part of the laboratory's data management system and can be used to flag unusual values of lead concentrations. Note that this probability model does not necessarily fit the data in the traditional statistical sense. That is, no effort has been made to examine the question, "Is the original distribution from which the data arose truly exponential?" Rather, this model has been selected for the purpose of decision-making and concentrating on the question, "Is this particular piece of data sufficiently unusual to be flagged?"

DISCUSSION

A principal advantage of this methodology is that the entire process can be carried out automatically by computer, producing high- and low-range limits as its output. Currently, the first phase of a general computer program to perform these steps has been completed, and a contract has been written to apply this program to selected water quality data files in Region V's Central Regional Laboratory in Chicago. The data screening range limits which result from this effort will be incorporated into the Region V Laboratory Data Management System (LDMS) first as part of the time shared, remote terminal system and later as part of the automated, minicomputer system to be installed at this laboratory.

Figure 1
Probability Density Function (PDF) of Weibull Distribution for Different Values of β, Where $\xi = 2.0$

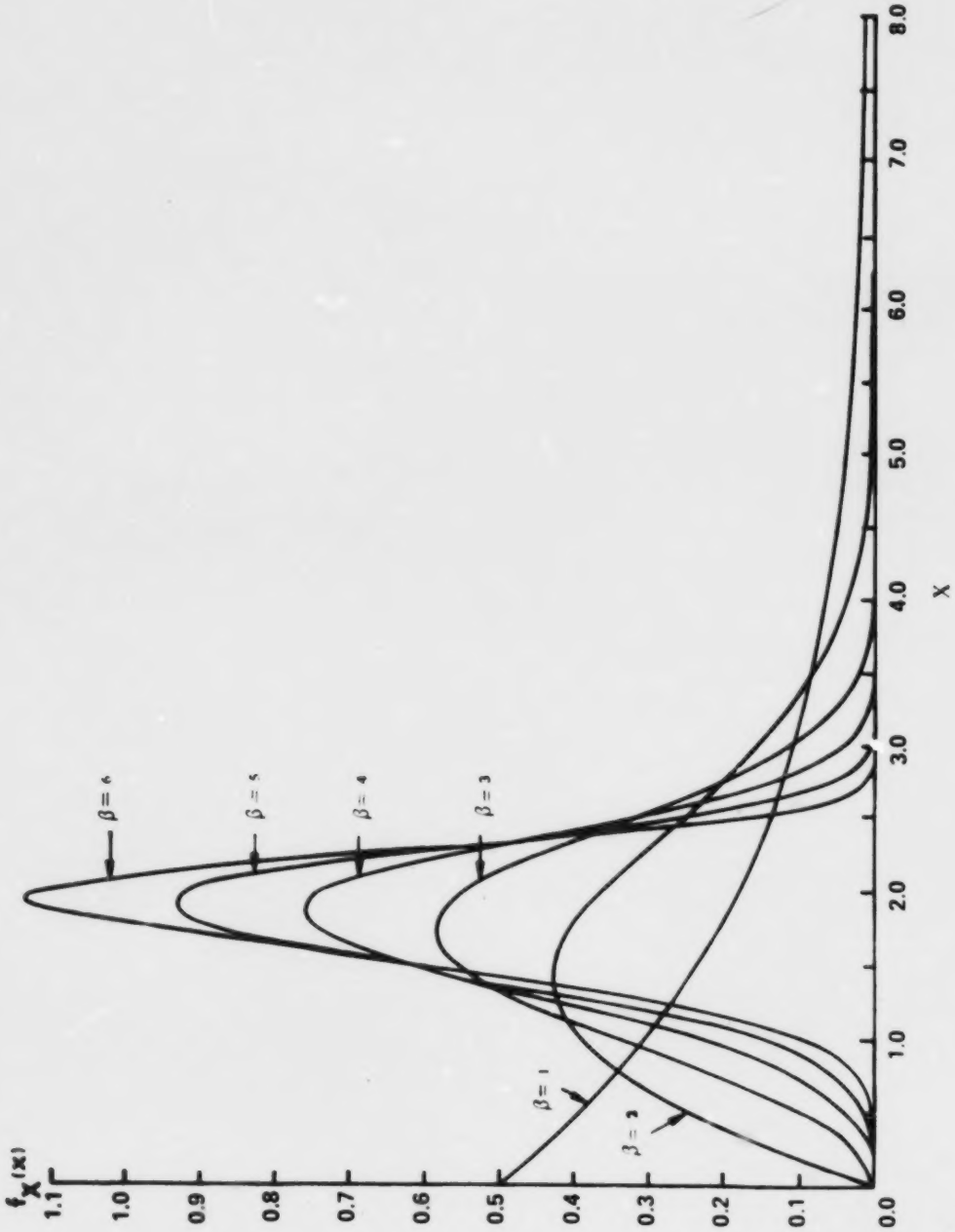
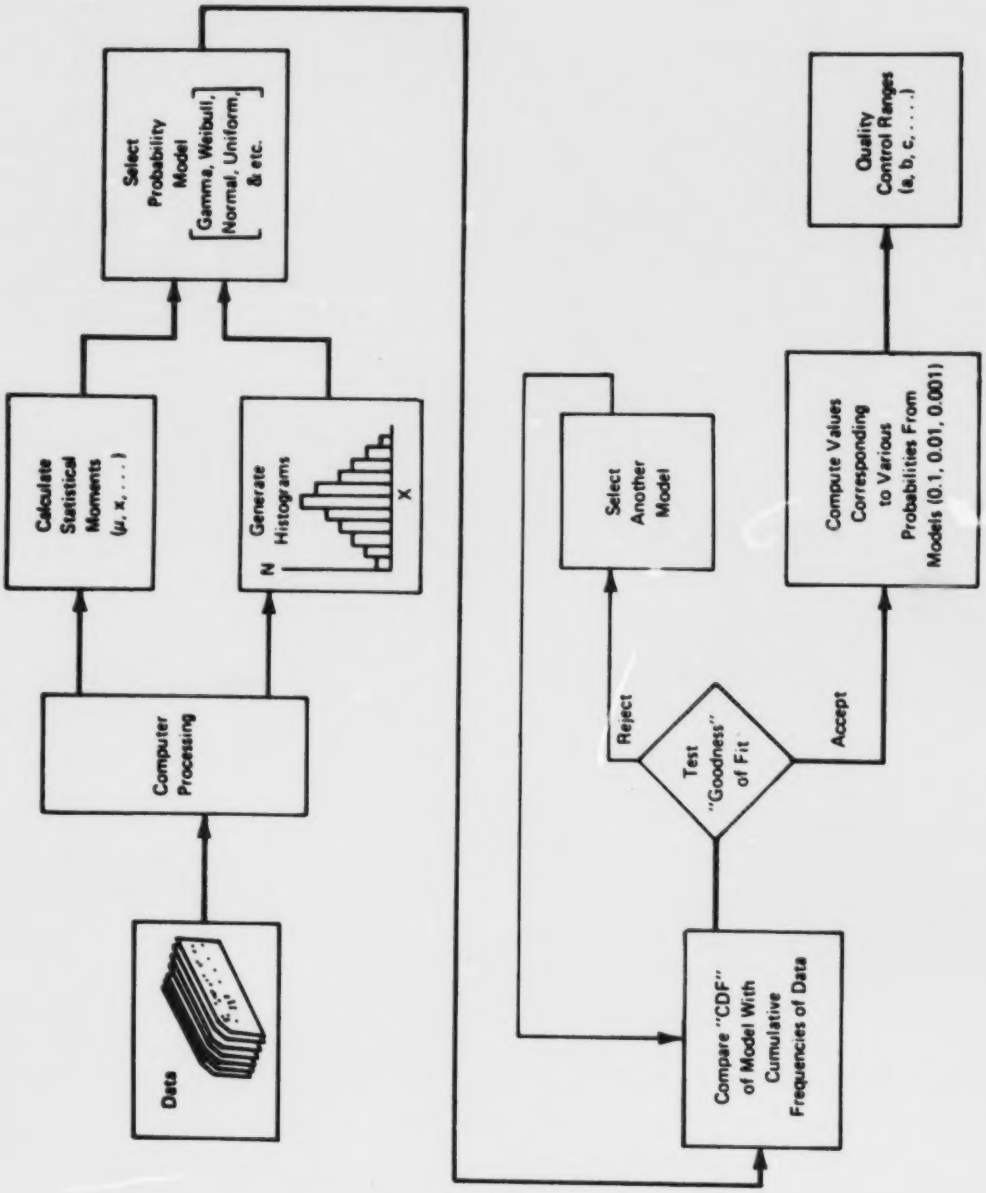


Figure 2
Steps Required for Probability Model Selection
In Determining Quality Range Limits



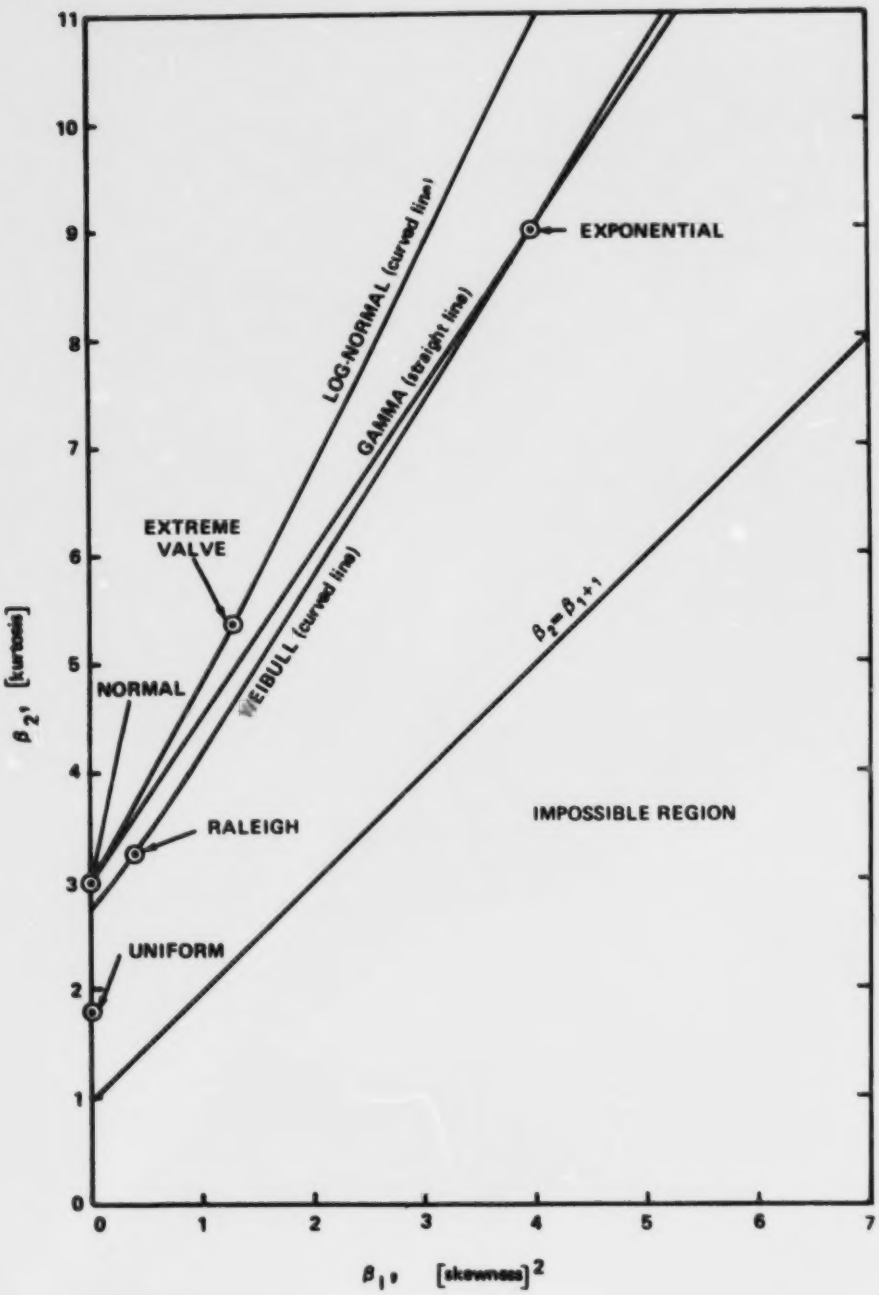


Figure 3
Relative Magnitude of β_1 and β_2 for Different Probability Models

```
                    * * *   DISTRIBUTION OF DATA (INTERVAL WIDTH =  0.2000)   * * *
TOTAL NO. OF VALUES = 4411    MEAN =   0.862002    VARIANCE =      1.09539986     STD. DEV. =  1.046614

NO. OF VALUES LESS THAN 0.0    =     0

   INTERVAL       NUMBER    PERCENT    +---------------------HISTOGRAM--------------------+    CUM. FREQ.
0.0   --0.2000    1012      22.943%    I************************                          I     22.943%
0.2000--0.4000     659      14.940%    I****************                                  I     37.883%
0.4000--0.6000     665      15.076%    I****************                                  I     52.958%
0.6000--0.8000     452      10,247%    I***********                                       I     63.206%
0.8000--1.0000     361       8.184%    I********                                          I     71.390%
1.0000--1.2000     309       7.005%    I*******                                           I     78.395%
1.2000--1.4000     213       4.829%    I*****                                             I     83.224%
1.4000--1.6000     139       3.151%    I***                                               I     86.375%
1.6000--1.8000     100       2.267%    I**                                                I     88.642%
1.8000--2.0000      52       1.179%    I*                                                 I     89.821%
2.0000--2.2000     106       2.403%    I**                                                I     92.224%
2.2000--2.4000      56       1.270%    I*                                                 I     93.493%
2.4000--2.6000      39       0.884%    I*                                                 I     94.378%
2.6000--2.8000      28       0.635%    I*                                                 I     95.012%
2.8000--3.0000      16       0.363%    I                                                  I     95.375%
3.0000--3.2000      39       0.884%    I*                                                 I     96.259%
3.2000--3.4000      24       0.544%    I*                                                 I     96.803%
3.4000--3.6000      22       0.499%    I                                                  I     97.302%
3.6000--3.8000      14       0.317%    I                                                  I     97.619%
3.8000--4.0000       5       0.113%    I                                                  I     97.733%
4.0000--4.2000      16       0.363%    I                                                  I     98.095%
4.2000--4.4000      11       0.249%    I                                                  I     98.345%
4.4000--4.6000       6       0.136%    I                                                  I     98.481%
4.6000--4.8000       4       0.091%    I                                                  I     98.572%
4.8000--5.0000       3       0.068%    I                                                  I     98.640%
5.0000--5.2000      12       0.272%    I                                                  I     98.912%
5.2000--5.4000       6       0.136%    I                                                  I     99.048%
5.4000--5.6000       5       0.113%    I                                                  I     99.161%
5.6000--5.8000       2       0.045%    I                                                  I     99.206%
5.8000--6.0000       0       0.0  %    I                                                  I     99.206%
6.0000--6.2000       2       0.045%    I                                                  I     99.252%
6.2000--6.4000       5       0.113%    I                                                  I     99.365%
6.4000--6.6000       7       0.159%    I                                                  I     99.524%
6.6000--6.8000       2       0.045%    I                                                  I     99.569%
6.8000--7.0000       1       0.023%    I                                                  I     99.592%
7.0000--7.2000       5       0.113%    I                                                  I     99.705%
7.2000--7.4000       1       0.023%    I                                                  I     99.728%
7.4000--7.6000       4       0.091%    I                                                  I     99.818%
7.6000--7.8000       2       0.045%    I                                                  I     99.864%
7.8000--8.0000       0       0.0  %    I                                                  I     99.864%
8.0000--8.2000       3       0.068%    I                                                  I     99.932%
8.2000--8.4000       1       0.023%    I                                                  I     99.954%
8.4000--8.6000       0       0.0  %    I                                                  I     99.954%
8.6000--8.8000       0       0.0  %    I                                                  I     99.954%
8.8000--9.0000       1       0.023%    I                                                  I     99.977%
9.0000--9.2000       0       0.0  %    I                                                  I     99.977%
9.2000--9.4000       0       0.0  %    I                                                  I     99.977%
9.4000--9.6000       1       0.023%    I                                                  I    100.000%
9.6000--9.8000       0       0.0  %    I                                                  I    100.000%
9.8000--******       0       0.0  %    I                                                  I    100.000%
                                       +-----------------------------------------------+

NO. OF VALUES EQUAL TO OR GREATER THAN 10.0000 =    0  ( 0.0  %)
```

Figure 4
Computer-Generated Histogram of 4411 Atmospheric 24-Hour Lead Concentrations Measured in U.S. Cities in 1968

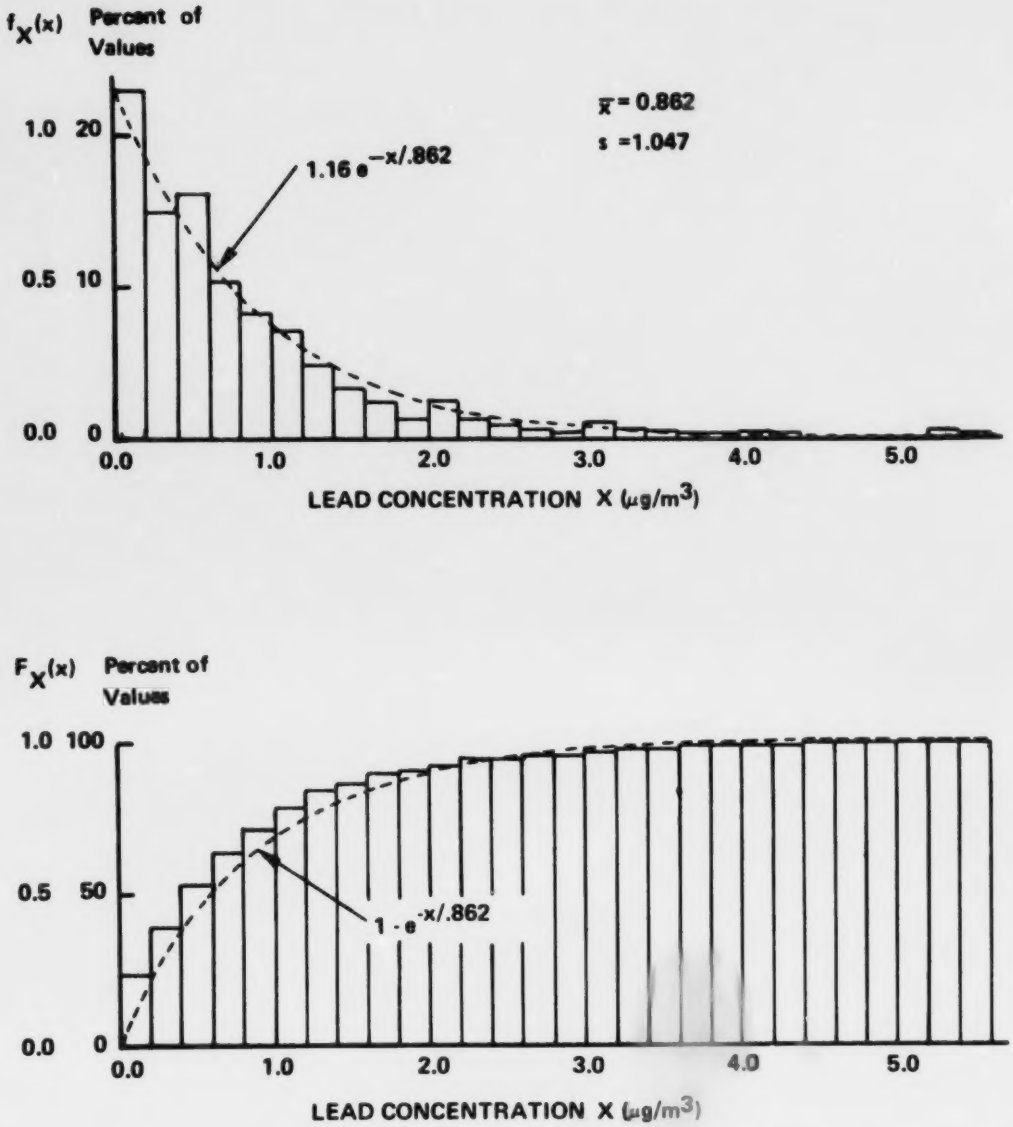


Figure 5
Distribution of 24-Hour Average Lead Concentrations From National Air A Sampling Network, 1968, Along With the PDF and CDF of the Exponential Probability Model

A SYSTEMATIC APPROACH TO WATER QUALITY TREND ANALYSIS

By Merlin H. Dipert* and Jon A. Abraytis

INTRODUCTION

A systematic approach using empirical models is proposed to investigate water quality data trends. A trend analysis procedure, with supporting computer programs, is presented.

For ease and economy of operations, three programs are used. One retrieves the data from STORET; another fits the data; and a third plots the theoretical curves, the points, and their confidence limits. In order to simplify the mathematics, the graph is considered as a quality control chart.

The function is assumed to be nonlinear in time; however, since the program takes its own derivative, the model can be changed by changing one card in the fitting program and one card in the plot program.

DISCUSSION

In investigating water quality trends, errors of measurement, which differ from random errors in the process, should be considered. In this paper, however, it is assumed that the points (individual points, replicated means, periodic means, or some similar values with their calculated errors) have been obtained by some standard method. This discussion is limited to the errors of fitting these points.

Consider the general n-dimensional mathematical model,

$$y = L(t) + F(t) + G(x) + S,$$

where L(t) is the linear trend line, F(t) is the nonlinear component of time, G(x) is a function of other parameters (e.g., latitude, ambient temperature, or depth of measurement), and S is the seasonal parameter in the same sense as used by Box-Jenkins.†

For preliminary investigations, the project staff selected a station in the Ohio River near Cairo, Illinois, and, using the specific model,

$$y = A + Bt + C \sin (D + Et),$$

and setting the period equal to one year, a 98 percent reduction of variance was obtained for temperature; a 28 percent reduction for dissolved oxygen; and a 30 percent reduction for chlorides.

Since the results are preliminary and there are no plots of dissolved oxygen and chlorides, and since the model should be changed, values are not given. The values obtained for centigrade temperatures, with their standard errors, are given in Table I.

Table I
Sample of Values Obtained From the Model

Parameters	Values
Intercept	16.1 + .31
Slope	-.00077 ± .00031
Amplitude	12 ± .23
Phase angle	-2.05 ± .02
Phase angle days	119 ± 1.1
Degrees of freedom	28
Variance	1.4
Reduction in Variance	98%

The theoretical curves, with the means, the 95 percent confidence limits lines for the monthly means, and the individual plots are presented in Figure 1.

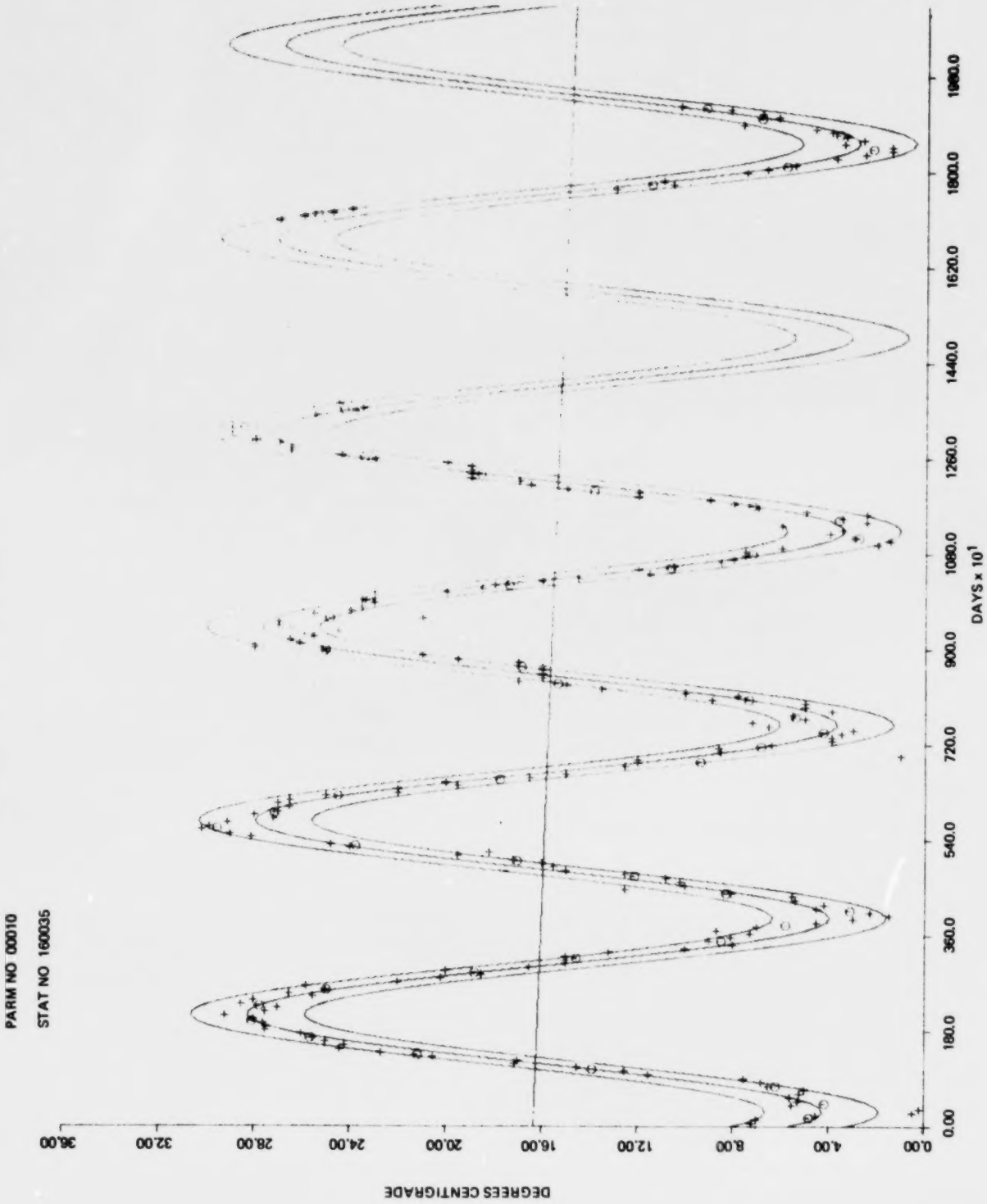
CONCLUSION

An approach which can be used to compare water quality parameters, determine linear trends over time, and calculate any of the parameters of the equations required by the physical scientist has been presented in this paper. When the mathematical models are the same, this approach will give the necessary transformation to make the curves coincide. Perhaps its most important use is to determine environmental quality trends (changes in parameters over time) to discover if enforcement policies have had a real effect on environmental quality rather than merely a "perceived" effect due to seasonal or other cyclical phenomena.

*Speaker

†G.P. Box and G.M. Jenkins, *Time Series Analysis, Forecasting, and Control* (San Francisco: Holden-Day, 1970).

Figure 1


PARM NO 00010
STAT NO 160035
DEGREES CENTIGRADE
0.00
4.00
8.00
12.00
16.00
20.00
24.00
28.00
32.00
36.00
DAYS x 10^1
0.0
180.0
360.0
540.0
720.0
900.0
1080.0
1260.0
1440.0
1620.0
1800.0
1980.0

SAMPLE FILE CONTROL FOR THE AUTOMATED LABORATORY

By Roman I. Bystroff

INTRODUCTION

During this year, a team at Lawrence Livermore Laboratory has engaged in feasibility studies for a number of water quality related laboratories in the Environmental Protection Agency (EPA). All of these laboratories are participating in the Cincinnati Pilot Project, an effort directed at developing minicomputer-based laboratory automation, which will be transportable to other EPA laboratories.

The team has definite ideas about laboratory automation, and it seemed sensible and proper from the inception that once sample data were in the computer, some form of sample file control would be used. Ideas about sample file control have since evolved, particularly as the regional laboratories emphasized its importance and benefits.

Today functional specifications for a sample file controller are nearing completion; this should fit the needs of laboratories with quite diverse operations, namely Region V and National Field Investigation Center (NFIC), Cincinnati. This paper will first describe the functional characteristics that are proposed to be implemented and also the value judgments inherent in the implementation.

FUNCTIONS OF THE SAMPLE FILE CONTROLLER

The central file in Figure 1 is referred to as the active status file (ASF) since it will contain all information about samples on work in progress. A laboratory supervisor uses this file as a log-book. On receiving information about a sample or related samples by groups, he enters this information and also indicates possible procedures, e.g., the test procedures. The program modules LOGIN and EDIT create these entries and aid by prompting him for the required information. Blanks are left for late information and filled in later.

Once samples are logged-in, an operator at some work station queries the file and selects those samples which he can handle and which are intended for him. The WORK SCHEDULE program module aids him in this function by passing all information required about the samples into his station file. The station controller is an applications program for the performance of an automated analysis procedure, including the essential quality control. The output of the station is one result for each sample test procedure, which is passed back to update the active status file. In addition, the station passes information about quality control statistics on each sample to another file, the primary data file (PDF). The file is archival; this information is required only when legal cases are called or for quality control auditing. Another archival file is the completed sample file (CSF), which can be accessed for laboratory statistics, cross-references to the primary data, and so forth. The function of the LOGOUT program module is to aid in the transfer of sample data from an "active" to a "completed" archival file.

A report generator module aids the operator by producing a listing of any specified data items or by producing standard formatted reports from any logically structured files, that is, the ASF, CSF and PDF.

The final functional module is a STORET converter which prepares the sample data stream in a format suitable for transmission to a national data system.

THE LOGICAL STRUCTURE OF FILES

In Figure 2, a tree structure with up to three levels of branches defines the organization of any file in the sample file controller (which does not include station files). The file is structured to reduce redundancy. On the branches are data items associated with descriptors. The descriptors and their meanings, as employed in the water samples active status file at Region V, are given in Table I. Note in Figure 2 that there are as many data items associated with one descriptor as there are branches at any one level. For example, at level 2, there might be six sample numbers which have the common descriptor (DESC 3) SAMPLE, and exactly six sample type items with the descriptor SAMPLETYPE (DESC 4). This somewhat minimal restraint on the allowed structures of the tree simplifies a structure specification. The presence of structure code permits a rather straightforward means of restructuring the files. A PDF with a different structure than the water ASF is shown in Figure 3.

Table I
Examples of Water Sample
ASF Descriptors

Descriptor		Description of the Data
Complete	Abbreviations	
Level 1		
OFFICE	OFF	District office
GROUP	GP	Study number
STUDY	–	Study description
GP STATION	GPST	Station number of group
DUE DATE	DUE	Group due date
SOURCE	SOR	Identification of river, lake, sewage
GP TAKEN	GPTAK	Group date taken
REC DATE	RECDAT	Group date of arrival
NUMSAM	NSAM	Number of samples
NUMPARM	NPARM	Number determinations to be performed
ACCT		Study account number
POLLUTE	POL	Group pollution index for ...
GROUP DATE	GPDAT	Date group logged out
DIRECTOR	DIR	Director of field office
STRUCTURE	STRUCT	Type of data structure
Level 2		
CRENUM	CRE	CRES sample number
SAMPLE	#	STORET sample #
SAMLOC	SLOC	Code for location where sample taken
TST DATE	TSTDAT	Date sample taken
SAM STATUS	SSTAT	Status of the sample
SAM ID	SSTAT	Sample identification
SAM TYPE	STYP	Type of sample
CUSTODY	CUS	Custody sample
SAM LAT	SLAT	Latitude of sample location
SAM LONG	SLON	Longitude of sample location
COMPOSITE	COM	Is the sample a composite
SAM TIME	STIM	Time sample taken
SAM WT	SWT	Sample weight (grams)
Level 3		
ANALYSIS	ANAL	Any parameter to be determined
VALUE	VAL	Concentration or value of a determination (see list of analyses for units)
METHOD	METH	Method of analysis
OPERATOR	OPER	Operator performing analysis
ANAL DATE	ANDAT	Date analysis finished
STATUS	STAT	Status of analysis
STORET	STOR	STORET parameter number

The use of a structure code is a way of satisfying the functional requirement that the user has the option to redefine what items he requires and their relationship to each other in the file controller. The importance of this in a pilot project and in the "tunability" of an implementation for diverse laboratories is that difficult reprogramming is avoided. The laboratories SFC needs are evolving and will continue to evolve. For example, at Central Research Laboratory, Region V (CRL-V), three structures (ASF types) are required for sample files because water samples, bubbler samples and high-volume air filter samples have different numbers and kinds of descriptors. If biological samples were also included under the computer SFC, another type of active status file could be added.

The water ASF in Figure 2 and PDF in Figure 3 illustrate the use of different keys in structured files. Samples are effectively organized by PROJECT or GROUP and are often kept together in batches for an analysis work session. The quality control data are keyed by the time-of-completion because analysis procedures vary in length. Various sample determinations are related to a calibration series (represented by coefficients of a least squares fit and the standard deviation in Figure 3). Retrieval of the primary data for a group of samples will be fastest if the cross-indexed time-of-completions are first found in the CSF under its key: GROUP. This then is an example of a trade-off: data are logically organized to be easy to input without the overhead of reordering data.

INTERACTIVE FEATURES OF THE SAMPLE FILE CONTROLLER (SFC)

The user's acceptance of a computer is very high on the list of functional values. After all, the laboratory chemist is not very burdened by choosing where to write his data on a standard form sheet. Program aids to log or request data should be functionally equivalent to those of the form sheet. This means memory aids (prompts), defaults to normal requests, and the freedom to enter, not enter or ditto data entries.

The LOGIN program module illustrates these kinds of features. A person logging in a group of samples may take various paths as he fills in the sample information in an ASF form sheet. If the SAMPLE TYPE happens to be the same for all the samples in this group, he merely types the description and ";ALL". If they differ, LOGIN will prompt him for the entry for each individual sample. At any time he can pass over an entry. It may be a delayed piece of information needed only for a final report, or one not always required. The LOGIN module can be used again to fill in previously unspecified data. In this case, the prompting would begin with the first blank in the form record and proceed from there. LOGIN would not be used to change existing items.

When the operator wants the SFC to "Do Something!" he must specify a command and a specification in a one-line statement. The commands are single words and the specification must answer "to what item or items?" in an already specified file. Example command lines are listed in Table II; they illustrate the form of the specification for the various program modules. Note that in line 1 the item "BOD5" is to be

Table II
Examples of Command Lines in the Sample File Controller Program Modules

Module	Command	Specifier or Output
EDIT		
	ADD	TEST BODS TO GROUP 101 SAMPLE 1 THRU 10, 12
	DELETE	TEST Hg FROM GROUP 101
	CHANGE	SAMLOC SAMPLE 10 THRU 20 TO 10.52W
	CHANGE	SAMLOC SAMPLE 10 THRU 20
		SAMPLE
		10"
RPG		
	LIST	GROUP 101 SAMPLE CUSTODY OPERATOR
	PRINT	GROUP 101 SAMPLE CUSTODY OPERATOR
		1 STA 10 JWB
		2 STA 10 JWB
	SUMMARY	DIVISION INDO GROUP 101
WORK SCHEDULE		
	FIND	Hg GROUP 101
	PRINT	
	SELECT	10 THRU 20, 22 Hg
	RESTORE	

added under the descriptor TEST for a specific set of samples. The delimitor ";" is used to mean the logical 'AND" of the two specifications GROUP:101 and SAMPLE:1 THRU 10,12. The forms ADD.. TO, DELETE ... FROM, CHANGE.. TO are expected. If the preposition is left out, it is assumed the operator wants to be prompted by sample number, so that he may enter individual values as in line 4. Note that if a descriptor (SAMLOC) is left unmodified, it is assumed to mean ALL items described by SAMLOC.

The report generator commands are such that anything in the file can be listed. The called-for level 1 descriptors will appear in a row at the top of the report, and level 2 and 3 descriptors will head columns of items across the page. The order of appearance will be the order in the command line. The example shown calls for all samples and all custody and all operators. The list would be shortened if, for example, the operator was specified OPERATOR:JWD. Special standard forms would be called by a more concise SUMMARY command. These could be samole summaries or managerial reports with column totals, if desired. The WORK SCHEDULE module commands are more terse because it is assumed that the only items of interest to specify are the ANALYSIS and the SAMPLE numbers. Thus the descriptors are left out, and only the item is stated.

CAPABILITIES FOR OFF-LINE DATA ENTRY

An SFC is most cost effective if parallel records are altogether eliminated. All the test results, including those which come from nonautomated tests (those used less frequently, presumably), can and should be entered into the computer files. There are two ways for accomplishing this. The EDIT program can update a sample log with a keyed-in result (see Figure 1). A better procedure would be for the user to write an applications program in the BASIC language to prompt him for the data in a systematic fashion and incorporate quality control. This has the added advantage that key-in errors can be reduced by checking against limits.

SUGGESTIONS FOR THE EFFECTIVE USE OF SAMPLE FILE CONTROL

In the EPA there has been some thought given to making use of preknowledge about samples in a passive sense. For example, Wayne Ott and W. Fairless have considered limits that results might be compared against. One use of such limits would be as a coarse filter for key-in errors (e.g., pH of > 14 is impossible for water).

Geographic keying of lake and river improbably high (or low) results may prove of value in some regions if sampling and seasonal variations are not serious. The SFC can accommodate these files, and the WORK SCHEDULE program could make these TEST-related parameters available to the analyst for purposes of double checking. It is suggested that in the compliance monitoring (C-M) program, license limits are quite specific and can be keyed by industry or industry-type and test procedure. In the passive use of these limits, a bell could be made to ring if a limit were exceeded. This may be unnecessary, but if one understands that automatic sample file control and lab automation compress the operational time scale for the laboratory, then these C-M limits could be used for closed loop options. For example, the reduction of the number of determinations in the laboratory could be achieved if a minimal quality control scheme were employed when the result (plus uncertainty) is less than the C-M limit. If a C-M limit (either a peak-allowable or maximum daily allowable) were reached, a conservative quality control scheme would be employed. After all, once the STORET baselines have been established, normal values are of incidental interest. The enforcement lawyers would have

to address this question, but it seems that this scheme is not a bias of results, but rather like a sliding uncertainty criterion.

Other examples of closed loop effects of automation on laboratory operations are not entirely related to the subject of this paper, but in response to queries from National Environmental Research Center (NERC), Corvallis, a random access auto sampling wheel is employed. It makes possible the scheduling of, and access to, samples to be rerun in the same work session.

The laboratories in question employ custody procedures. The presence of status conditions and custody identification in the SFC files can be made effective as a means of locating the physical sample and possessor. Similarly, the analyst can be encouraged to conform to a strict protocol by the SFC.

Management reports were not a high-priority function of the SFC design, but are a bonus which can help operational evaluations.

DESIGN VALUES FOR A SAMPLE FILE CONTROLLER

At the stage of thinking about the pilot project laboratory as a system (that is, in terms of what is done and what flexibilities exist that should be preserved when automated), value judgments must be made as to what features are critical for the success of the system as a whole.

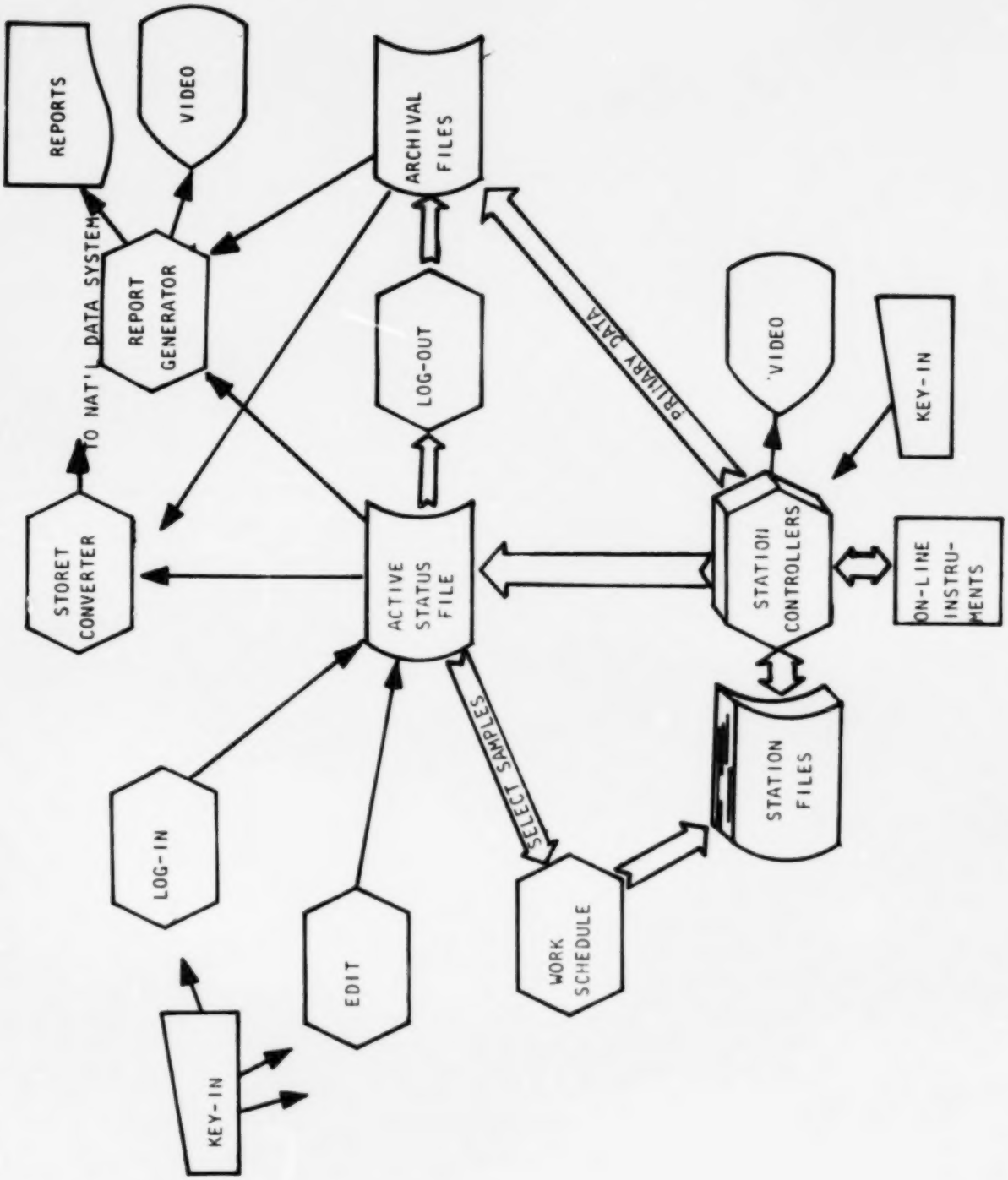
Some of these may be evident from the preceding description. The data flow must correspond to existing operations. This includes a query and input capability for the engineers, chemists and managers; therefore, the accesses to the data base must be interactive and easy to u

There are staffing constraints in the laboratories. A choice is made to allow for the problem-oriented chemist to engage in the solution to laboratory procedure problems. In the automation of the laboratory, the choice of an easy-to-use BASIC language will allow the user to be involved in the growth of the system applications. Sample file control has a lesser, but still evident, requirement to be adaptable to evolving needs. Value is placed on reprogrammability by the user. This is particularly valid in a pilot project in which options are to be kept open until tested and evaluated for usefulness, user acceptance, cost effectiveness, integrity, and so forth.

The adaptability of the sample file control is an important value which addresses the objective of transportability to other EPA laboratories. A modular design addresses this. The laboratory automation programs and the sample file control routines are separate modules in the whole system. The options remain open for resident versus hierarchal (CPU-CPU) implementations.

For the present it is believed that the Data General-supported BASIC language can be used to write the SFC in a way that will satisfy the values mentioned so far. The NOVA-840 has the capability in its foreground/background disc operating system. However, the response time of such an implementation is still a question. The job is to find the right trade-offs to implement the functional values mentioned in this paper. When this is done, the prime objective will have been achieved: to have a cost effective laboratory-based system in which the user is given authority for his portion of the responsibility, that is, to efficiently produce quality analytical results.

Figure 1
Data Flow Within the Sample File Controller



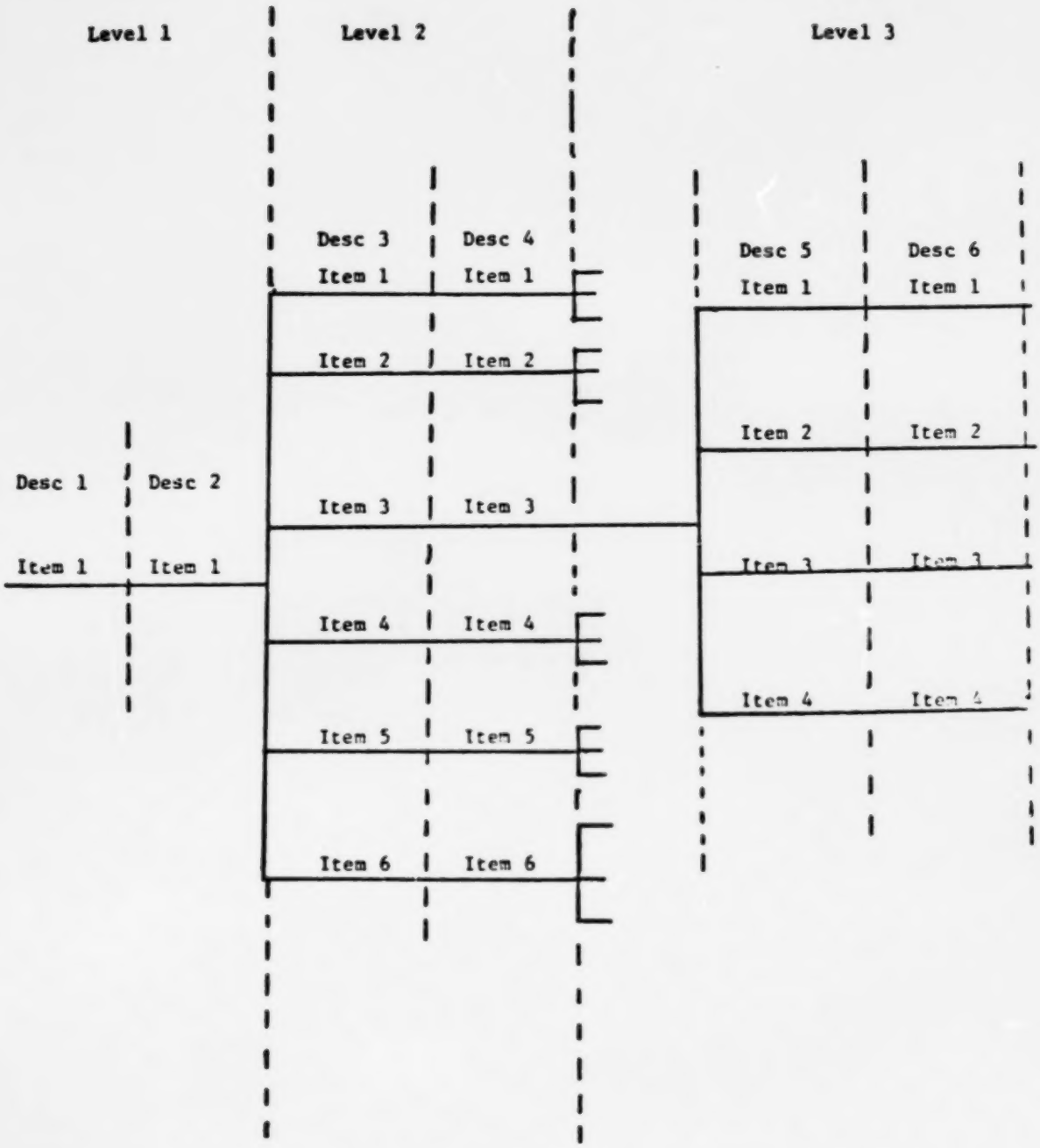


Figure 2
General Data Structure

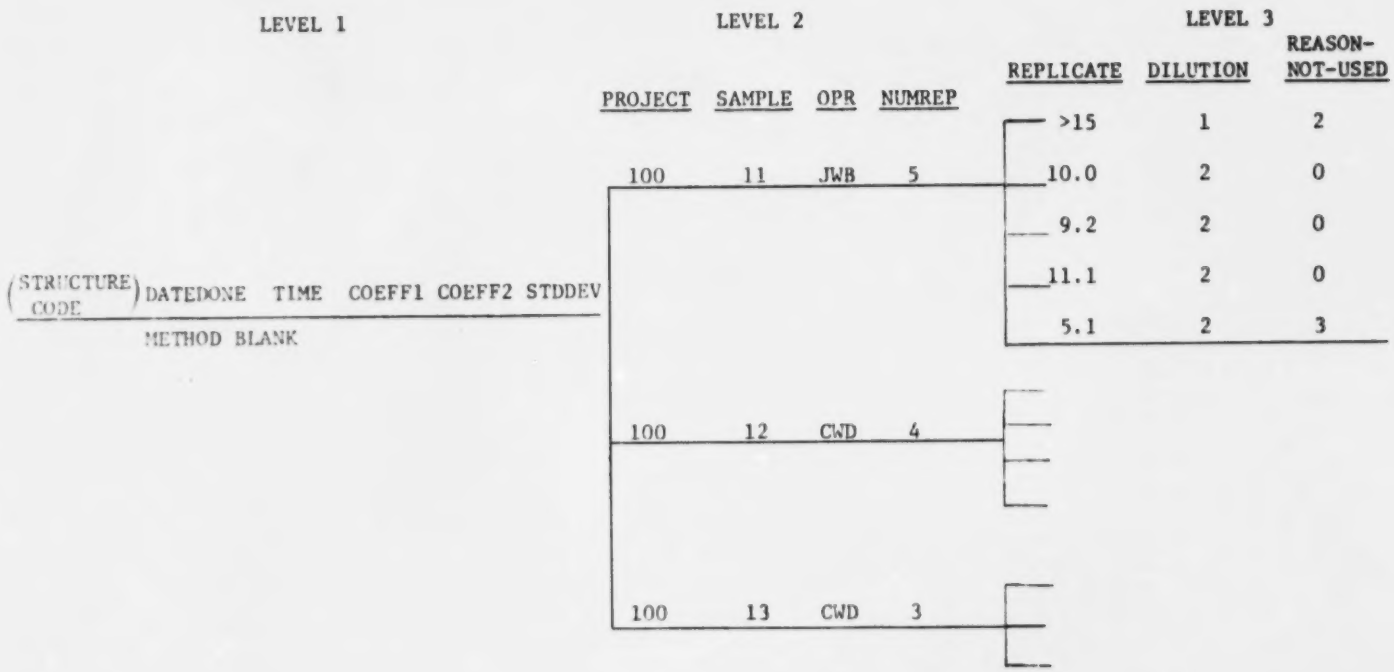


Figure 3
Logical Structure of a Primary Data File for Quality Control Documentation

A DATA REDUCTION SYSTEM FOR AN AUTOMATIC COLORIMETER

By K.V. Byram,* F.A. Roberts, and L.A. Wilson

INTRODUCTION

The Consolidated Laboratory Services Program (CLS) at the Pacific Northwest Environmental Research Laboratory (PNERL) has analyzed a large variety of samples supporting the research programs and, until recently, the EPA Region X programs. It has been found useful to employ the Technicon Autoanalyzer II for performing analyses, where the volume of work to be done justified the automated system. At PNERL, the Technicon Automated System has been averaging about 3000 determinations per week, running from three to five channels simultaneously.

At this rate, it seemed imperative to automate the data capture process for information coming from these instruments if their full potential to reduce manpower requirements was to be utilized. Although Technicon, Inc. produces both a printer (producing output on adding machine width tape) and a teletypewriter (producing full width or punched paper tape), these outputs are of only minor advantage if a full complement of analytical quality control procedures is employed.

WHAT THE TECHNICON PRODUCES

The Technicon autoanalyzer is a continuous flow reagent-mixing and colorimeter system which selects the sample to be analyzed from a series of cups. The liquid from each cup is followed by a wash to discriminate samples. The cups are selected at rates of around 50 per hour. After a uniform delay for reagent mixing and color development, a voltage is produced from the colorimeter. The voltage is proportional to the amount of constituent in the cup, which is a peak above the zero constituent in the wash (Figure 1).

At PNERL, the minimum run consists of a series of standards, a series of unknowns (samples whose concentration is unknown), and a series of standards. If the run is extended, another unknown/standard series follows the first. The "standards" consist of initial blanks (distilled water) to clean out the system, followed by a series of at least eight standard solutions spread over the range of interest, followed by trailing blanks. The "unknowns" include unknowns, replicates of knowns, and unknowns to which a known has been added (spikes). The normal run consists of about 20 standards in two groups of ten, and 105 samples, which, when combined and interspersed with blanks, spikes, and so forth, comes to a total of 156 peaks. It is this series which is presented to the computer programs for reduction.

NECESSITY FOR AUTOMATING DATA CAPTURE

The primary reason for the decision to process the autoanalyzer output by computer was the volume of data expected. It is a reasonable job to reduce the output from the instrument run in production for an hour or two a day, spending an equal time reporting the data. But keeping up with the instrument running in production for 15 hours a day was another matter.

In convincing the staff that a more fully automated system would be useful, considerations such as the availability of a programming team with experience in this kind of work and of a predecessor program to perform parts of the work weighed heavily. Not least important was that many of the simplifying assumptions on which Technicon, Inc., markets equipment for water analyses were not valid for this particular application. Since the colorimeter output is produced directly in absorbance units, it is assumed that the concentration can be read directly from the chart or printout with automatic blank correction. The assumptions were not reliable enough for routine use at the levels and precision needed.

PROCESSING THE STANDARDS

When running colorimeteric analyses manually, it is normal procedure to run a set of standard solutions and plot their concentrations and absorbances on graph paper. Using the chemist's best judgment, a straight line is then drawn as closely as possible through the points; using the standard curve, the analyst estimates the concentration of an unknown from its absorbance. The difficulty in automating this process is in duplicating the chemist's judgment. If concentration is plotted on the x-axis and absorbance on the y-axis, it is common, for example, for the curve to flatten out at the high end due to deviation from Beer's Law. If most of the unknowns are in the high concentration range, the system is reconfigured so that it is linear for the region of interest. If

*Speaker

the values for the unknowns fall in the linear region, the higher standards are ignored in drawing the line. Another problem occurs when one of the standard solutions is inaccurate because of aging, error in preparation, or contamination in handling. Knowing that this possibility exists, the analyst simply ignores that point when drawing the line.

The immediately apparent way to automate this process is to use a least squares curve fitting routine (linear regression). This technique, however, tends to minimize the deviation of the standards from the line equally at any level. If a typical deviation of a series of standards such as .01, .02, .05, .1, .2, .5 is .015, the least squares routine will place the line through the points so that the .015 deviation is typical at the .01 level, as well as at the .5 level. This results in a very accurate curve near the high end and a very imprecise one near the low end.

The staff has used a modified least squares technique which seems to produce a "chemist compatible" curve most of the time. It is based on the chemist's specification of two quantities: the minimum value detectable with the system (herein called "precision") and the percent accuracy possible with near full-scale values. If the precision is .01 and the accuracy 5 percent, for example, the chemist reports his result, as do manufacturers of instruments such as voltmeters, as "within .01 mg/l or 5 percent, whichever is greater." It turns out that below a mg/l value of the precision divided by the accuracy, which is .01/.05 or .2 in this case, the precision will be greater, while above that, the accuracy will be greater (because 5 percent of .2 is .01).

The curve-fitting routine is an iterative process which (I) uses a conventional technique to compute an equation, $c = mx+b$, where x is the absorbance and c is the concentration. Next (A) the value of m so obtained is used to compute a new value of b using only the standards below the precision-accuracy ratio. Then, (B) that value (essentially, the baseline) is subtracted from each of the standards above the precision accuracy ratio to obtain a new value of m. Steps A and B are then performed over and over again until the changes in either b or m are trivial (0.1 percent) from iteration to iteration.

Once a trial value of m and b are obtained with this method, the predicted concentration of each standard, using the equation, is compared with its known concentration, computing a percent and mg/l deviation for each one. (II) If any of the low standards have a greater mg deviation than the precision, the worst one is deleted from the set. If any of the high standards have a greater percentage deviation than the accuracy, the worst one of these is deleted also. The iterative procedure (IA and IB) in the previous paragraph is then performed to yield a new value of b and m. These two steps (I and II) are repeated until the percentage and mg/l accuracy and precision are satisfied, or until too many standards are being removed.

A set of standards both precedes and follows each set of unknowns, and the slope and intercept to compute a concentration for each unknown is linearly interpolated from both sets of standards. An example output, showing the final computation for each set of standards and a portion of the computations for the unknown is shown in Figure 2.

REJECTING ERRONEOUS RESULTS

Inaccurate Standards

If a satisfactory curve cannot be fitted to either set of standards surrounding a sample batch, no computations for unknowns between those sets are made. They are resubmitted for analysis at the same dilution factor unless a particular unknown peak was off scale, suggesting a greater dilution factor.

Shoulder or Off-Scale Peaks

We must turn to the hydraulics of the system to understand another reason for rejects. The system pumps the sample through its inlet tube continuously; this tube is moved from sample cup to wash to next sample cup, and so on. The wash is necessary to clean out the system so that the constituent in one cup does not carry over to the next and bias its results. In the case of an unusually strong unknown followed by a weak one, the wash may not be complete enough. In the system, a peak which is less than 10 percent of the previous one is rejected and resubmitted for analysis because of this problem. When an unknown is strong enough to yield an off-scale reading, the next two following peaks are rejected. The off-scale peak is resubmitted at a high dilution factor.

Negative Results

If a peak falls below the lowest b value from the standards computation, interference is suspected, especially if it falls far below it. In our case, if the negative concentration is greater in absolute value than the precision, the unknown is resubmitted for analysis.

Standards Extrapolation

If the peak of a given unknown is higher than the highest standard peak, the unknown must be resubmitted for analysis at a higher dilution factor. This procedure is required since the apparent concentration is greater than the highest standard, and extrapolation from the curve developed for the concentration being observed may produce invalid results.

Quality Control

At regular intervals throughout a set of unknown samples, the following set of quality control solutions is interspersed: a wash, a known solution, a wash, an unknown, a repeat of the unknown, and the unknown spiked with the known solution. The spiking is such that the concentration measured in the spiked unknown should be equal to the concentration of unknown plus the concentration of the known which was added to it. It is required that the measured concentration of the spiked unknown be within 10 percent (in most cases) of the calculated concentration. Since there are two replicate analyses of the unknown alone, the spike must be within 10 percent of calculation using either replicate. If it is not within 10 percent, the sample is resubmitted for analysis.

Comments

Certain one-letter comments are enterable by the analyst when the computations are made. Most such comments refer to the original disposition of the unknown solution, such as M for missing, or X for improper sample preservation, etc. These comments cause the program not to calculate a value for the sample because it was unavailable when the analysis was performed. Instead, the sample is reported with no result, but with its comment. Another class of comment, such as U for unreliable, causes the program to resubmit the sample for analysis. This might be used when the chemist notes something irregular in the peaks which is not detectable by the program.

NITRATE-NITRITE ANALYSIS

The nitrate analysis consists of a reduction of nitrate to nitrite with measurement as nitrite. Any nitrite in the sample is also measured, so the final result is nitrate plus nitrite, expressed as mg/l of nitrogen. Although many laboratories assume that the nitrate and nitrite are additive, this is not the case in samples with high nitrite, such as a nitrifying sewage. The difficulty is that some of the nitrite ion becomes further reduced to ammonia and is not measured. The resulting "nitrate plus nitrite" value is not a simple sum, but is something like "nitrate plus part of the nitrite." To compute the nitrate from this value requires subtracting some fraction of a measured nitrite value, a fraction determined by running nitrite standard solutions through the reduction process. The analysis dictates, therefore, that the entire standards computations be carried out three times: once for nitrate alone, once for nitrite alone, and once for nitrite passed through the reduction process to obtain values for nitrate, nitrite, and of course, their sum.

EQUIPMENT CONFIGURATION

Technicon autoanalyzers used are Model AA-II, consisting of two sets of three channels each, with strip chart, 4-inch wide paper tape printed output, and teletype with print and paper tape output. The strip chart is a continuous record of the series of peaks which come from each sample cup (Figure 1). The printer and teletype outputs (Figure 3) are digitized values of the peaks, which are sampled at the same time that the sample cups are changed. Phasing (that is, making sure the peak is sampled at the top rather than the side) is a function of the flow-through time for the analysis. Since this is constant for any particular setup, it can be set by the operator at the beginning of a run. The phasing for each channel is independent, so that the peaks need not occur simultaneously in all channels. Each time a peak is sampled for printing, a mark is made on the continuous strip chart record so the result can be checked later if necessary.

In this system, the paper tape produced by each teletype is hand-carried daily to the Oregon State University (OSU) Computer Center. This is much less expensive than feeding the information directly to the computer by putting the teletype on line as it is printing from the autoanalyzer. Since the output rate is so low (about 2000 characters per hour) from the autoanalyzer, hand-carrying the tape to the high-speed reader at the computer center is cheaper than reading it in a teletype speed from the laboratory.

Once in the computer, the tape is processed interactively with the operator at a terminal having considerable flexibility in editing the paper tape. This is routinely useful if the chemists let the teletype run out of paper tape or if they are unable to find a sample. It is also useful in cases of power failure, inadvertent omission of samples, insertion of extra samples, and so forth

In addition to the computational difficulties encountered in obtaining some sort of computer compatible output directly from the autoanalyzer, there is a sample identification problem. It is necessary, of course, to assign both identification numbers to samples and concentrations to standards before the computer compatible absorbance values have meaning. If this is done manually, at least some of the advantages in doing the computations automatically is lost. The system ties in with a laboratory sample handling system (SHAVES) which existed before autoanalyzer computer programs were developed. One of the autoanalyzer Technicon system programs, LISTGEN, accesses the file that holds the SHAVES analytical requests and produces a list of samples to be analyzed, complete with replicates, spikes, and standard solutions (Figure 4). The analyst then uses the list to fill the sample cups which are input to the autoanalyzer. When the results are computed, a copy of the list which has been kept in the computer is used to identify the results. The regular repetition of replicates, spikes, blanks, and the descending values of peak heights for the standards makes it easy to be sure that the right identification goes with each peak.

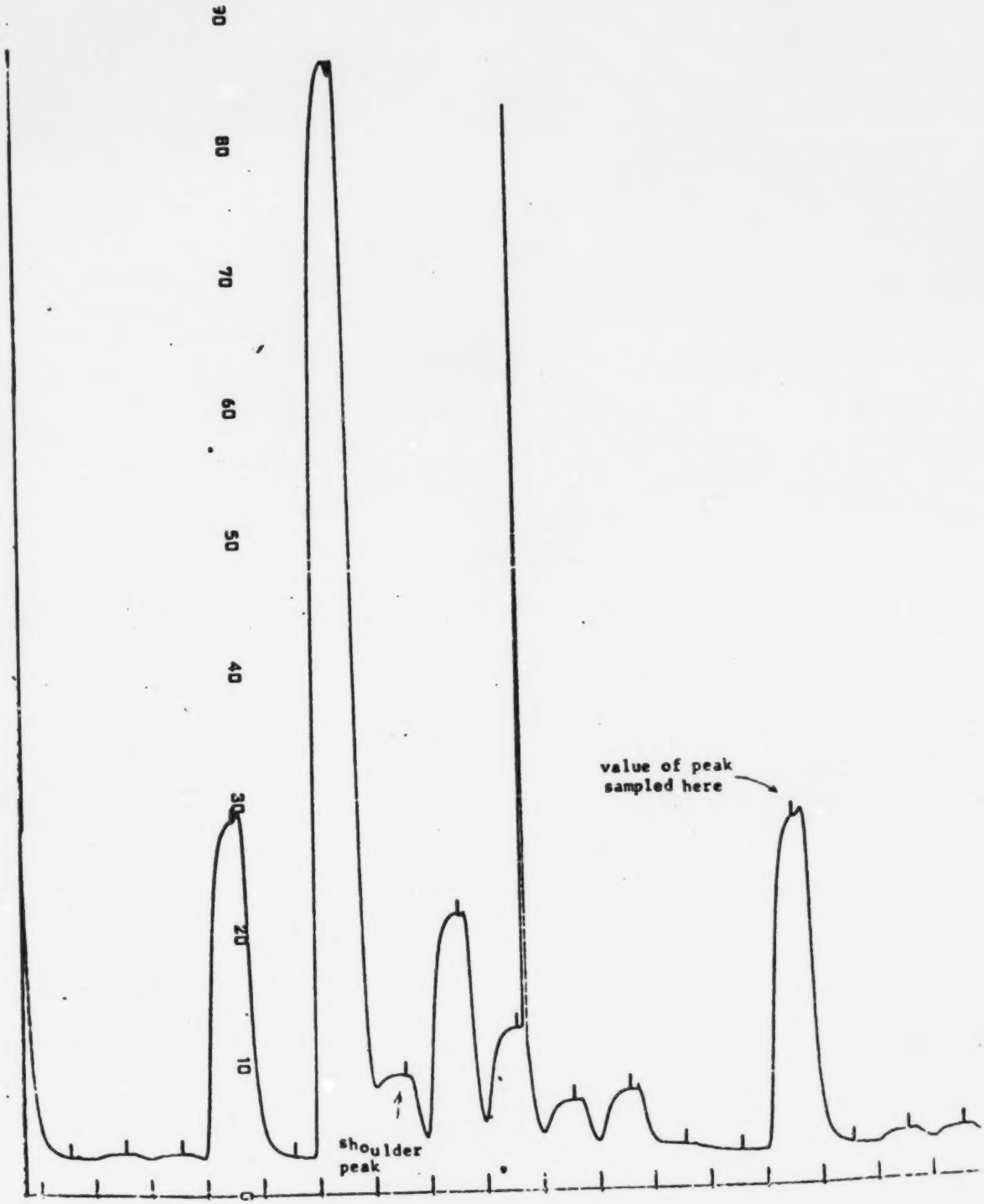


Figure 1
Strip Chart Recorder Output

```
BEGIN BATCH 9120817 FOR    625

ST    625  9/17/74    1510    (09/18/74  9120817)
C/A=       .6012

          CONC     PK HT      BASE      COA       DEV      %DEV
149     5.0000    8.7600     .4439     .6012         0         0
150*    2.0000    3.1300     .4439     .7446   -0.3850  -19.2490
151     1.0000    2.1400     .4439     .5896     .0198    1.9793
152      .8000    1.7300     .4439     .6220   -0.0267   -3.3395
153      .6000    1.4200     .4439     .6147   -0.0131   -2.1833
154      .4000    1.1600     .4439     .5585     .0306    7.6444
155      .2000     .7800     .4439     .5950     .0021    1.0533
156      .1000     .6200     .4439     .5677     .0059    5.9082
157      .0500     .5600     .4439     .4305     .0198   39.6676
160          0     .3800     .4439         0   -0.0384        -0


ST    625  9/17/74    1510    (09/18/74  9120817)
C/A=       .6063

          CONC     PK HT      BASE      COA       DEV      %DEV
240     5.0000    8.6800     .4330     .6063     .0000     .0000
241*    2.0000    3.1100     .4330     .7471   -0.3770  -18.8503
242     1.0000    2.1200     .4330     .5928     .0228    2.2772
243      .8000    1.7200     .4330     .6216   -0.0197   -2.4677
244      .6000    1.4200     .4330     .6079   -0.0016   -0.2712
245      .4000    1.1600     .4330     .5502     .0407   10.1847
246      .2000     .7400     .4330     .6516   -0.0139   -6.9504
247      .1000     .6200     .4330     .5349     .0133   13.3450
248      .0500     .5100     .4330     .6498   -0.0033   -6.6927
251          0     .3700     .4330         0   -0.0382        -0


SUMMARY FOR PARAM    625    9/17/74 1510    ENTERED  09/18/74 BATCH 9120817
 NO    LABNO     RESULT    RAW ANSWER    PKHT   BASE   FACT   REPORT RERUN RE
  1   5-25-  0         0 U(          0)    .50    .44      1     NO      ANALYS
  2   5-25-  0         0 U(          0)    .72    .44      1     NO      ANALYS
  3 SK   2.000     2.100  (     2.0698)   3.88    .44      1
(RECOVERY IS   2.0%          2.0000 PREDICTED)
  4 10-35-  0<      .100  (    -0.0021)    .44    .44      1     YES
  5 10-35-  1       .200  (      .1968)    .77    .44      1     YES
  6 10-35-  7<      .100  (      .0945)    .60    .44      1     YES
  7 10-35- 11<      .100  (      .0524)    .53    .44      1     YES
  8 29-31- 23<    10.000  (    -9.2058)    .29    .44    100     YES
  9 29-31- 28<      .100  (    -0.0859)    .30    .44      1     YES
 10 29-32-  8      9.900  (     9.8625)   2.08    .44     10     YES
 11 29-32- 17     26.000  (    25.8267)   1.30    .44     50     YES
 12 29-33- 37     33.000  (    33.0587)   1.54    .44     50     YES
 13 29-33- 37     32.000  (    31.5573)   1.49    .44     50     YES
 14 29-33- 38      1.400  (     1.3734)    .67    .44     10     YES
 15 29-33- 44     24.000  (    24.3383)   1.25    .44     50     YES
 16 29-35-  7      5.700  (     5.6710)    .63    .44     50     YES
 17 29-35- 10      6.300  (     6.2777)    .65    .44     50     YES
 18 29-35- 11     28.000  (    28.4808)   5.17    .44     10     YES
 19 29-35- 12      3.400  (     3.3654)   1.00    .44     10     YES
 20 29-35- 22<     5.000  (    -0.0336)    .44    .44     50     YES
 21 29-35- 23     14.000  (    13.9677)   2.76    .44     10     YES
 22 70-13-304       .400  (      .3669)   1.05    .44      1     YES
 23 70-13-304       .300  (      .3429)   1.01    .44      1     YES
 24 70-31-450       .800  (      .8188)   1.80    .44      1     YES
 25 70-32-602     18.000  (    17.5827)   1.90    .44     20     YES
 26 70-32-603       .900  (      .8973)   1.93    .44      1     YES
 27 70-32-604      2.600  (     2.6441)   4.83    .44      1     YES
 28 70-33-527       .200  (      .1867)    .75    .44      1     NO      SHLDR
 29 70-33-551<      .100  (      .0423)    .51    .44      1     YES
 30 70-33-552<      .100  (      .0785)    .57    .44      1     YES
 31 70-33-579      1.500  (     1.5674)    .96    .44      5     YES
 32 70-33-580      1.800  (     1.8374)   3.49    .44      1     YES
 33 70-33-580      1.800  (     1.8315)   3.48    .44      1     YES
```

Figure 2
Computer Output From Processing an Autoanalyzer Run

027.	71	0.31	25	0.65	65	0.61	-
028.	71	0.31	25	0.64	65	0.87	
029.	71	0.31	25	0.45	65	1.87	
030.	71	0.31	25	0.46	65	0.58	
031.	71	0.31	25	0.57	65	0.55	
032	71	0.31	25	1.00	65	1.90	
033	71	0.31	25	0.35	65	6.60	
034.	71	0.31	25	0.87	65	0.56	
035.	71	0.31	25	0.89	65	0.57	-
036	71	0.31	25	2.34	65	1.02	-
037.	71	0.31	25	0.40	65	2.41	-
038.	71	0.31	25	0.36	65	0.57	
039.	71	0.31	25	1.26	65	0.55	
040.	71	0.31	25	4.14	65	1.28	
041	71	0.31	25	0.37	65	3.77	-
042.	71	0.31	25	0.40	65	0.58	
043.	71	0.31	25	4.66	65	0.55	
044	71	0.31	25	0.23?	65	3.36	
045.	71	0.31	25	0.46	65	0.24?	
046.	71	0.31	25	0.40	65	0.62	
047.	71	0.31	25	0.97	65	0.55	
048.	71	0.31	25	2.04	65	0.69	-
049.	71	0.31	25	0.38	65	1.25	-
050.	71	0.31	25	0.40	65	0.56	
051.	71	0.31	25	1.35	65	0.61	
052	71	0.31	25	5.00	65	0.55	-
053	71	0.31	25	0.38	65	0.53	
054	71	0.31	25	0.38	65	0.55	-
055	71	0.31	25	0.35	65	0.53	
056	71	0.31	25	0.35	65	0.24?	-
057.	71	0.31	25	0.50	65	0.24?	E↑
058.	71	0.31	25	1.21	65	0.91	
059.	71	0.31	25	0.43	65	0.59	
060	71	0.31	25	0.41	65	0.66	
061.	71	0.31	25	0.69	65	0.84	B↓
062	71	0.31	25	0.42	65	0.54	
063.	71	0.31	25	0.67	65	0.55	
064.	71	0.31	25	1.31	65	0.88	
065.	71	0.31	25	0.40	65	2.18	
066.	71	0.31	25	0.38	65	0.56	
067.	71	0.31	25	0.44	65	0.56	
068.	71	0.31	25	0.61	65	0.59	
069.	71	0.31	25	0.37	65	0.72	-
070.	71	0.00?	25	0.38	65	0.53	
071.	71	0.32	25	1.26	65	0.52	
072	71	0.32	25	4.26	65	0.52	
073	71	0.32	25	0.38	65	0.59	
074.	71	0.32	25	0.47	65	0.57	
075	71	0.32	25	0.42	65	0.51	
076	71	0.32	25	0.55	65	4.21	-
077.	71	9.28	25	0.35	65	0.24?	
078	71	9.42	25	0.36	65	0.61	
079.	71	9.45	25	0.39	65	0.54	
080.	71	9.45	25	0.53	65	0.77	
081.	71	0.39	25	0.35	65	1.58	
082	71	0.43	25	0.38	65	0.54	
083	71	0.43	25	1.75	65	0.54	
084.	71	0.42	25	5.93	65	0.64	
085	71	0.42	25	0.37	65	1.08	

Figure 3
Example of Autoanalyzer Teletype
Output, as Annotated

"A" 16 Sep 74

BATCH 9120417 TRAY 1 PAGE 2

2045

(ADD ACID

TRAY NO	625 PRINTR LINE	665	LAB NUMBER	DILN FCTR	RUN FOR 1 2 3 4	RERUN CODE
1	163	156	05-25-000	1	. .X.	- -9-
2	4	157	-000*	1	. .X.	- -9-
3	5	8	SPIKE			
4	6	9	10-35-000	1	. .X.	- -4-
5	7	160	-001	1	. .X.	- -4-
6	8	1	-007	1	. .X.	- -6-
7	9	2	-011	1	. .X.	- -6-
8	170	163	29-31-023	100	. .X.	- -F-
9	171	164	-028	1	. .X.	- -F-
10	2		-32-008	10	. .X.	- -9-
11	3		-017	50	. .X.	- -F-
12	174	167	-33-037	50	. .X.	- -3-
13	5	8	-037*	50	. .X.	- -3-
14			-038	10	. .X.	- -4-
15			-044	50	. .X.	- -3-
16			-35-007	50	. .X.	- -3-
17			-010	50	. .X.	- -3-
18			-011	10	. .X.	- -4-
19			-012	10	. .X.	- -4-
20			-022	50	. .X.	- -3-
21			-023	10	. .X.	- -4-
22	184	177	70-13-304	1	. .X.	- -9-
23	5	8	-304*	1	. .X.	- -9-
24			-31-450	1	. .X.	- -4-
25			-32-602	20	. .X.	- -3-
26			-603	1	. .X.	- -4-
27			-604	1	. .X.	- -4-
28			-33-527	1	. .X.	- -8-
29			-551	1	. .X.	- -6-
30			-552	1	. .X.	- -6-
31			-579	5	. .X.	- -3-
32	194	187	-580	1	. .X.	- -4-
33	5	8	-580*	1	. .X.	- -4-
34	6	9	-581	1	. .X.	- -4-
35	7	190	-587	1	. .X.	- -8-
36	8	1	-34-541	5	. .X.	- -3-
37	199	192	BLANK	1	X.X.X.X	- - -
38	200	193	BLANK	1	X.X.X.X	- - -
~~39~~	201		05-35-001	1	.X.X.	-7-7-
~~40~~	202	195	-036	1	.X.X.	-7-7-

~~625~~
~~Change PK. HTS. Lines~~
~~0.77~~ ~~Not rec.~~
~~0.52~~

665
Change PK. HTS. LINE
163: 2.75
164: 0.73

Figure 4
Computer Prepared, Analyst Annotated, List of Analyses to Be Run

A CHEMICAL INFORMATION SYSTEM

By Stephen R. Heller

This presentation will discuss the ways in which the Agency is using the National Institutes of Health (NIH) Chemical Information System (CIS) and ways in which the CIS can be modified and/or expanded for further Agency use. (Figure 1)

Figure 2 is a list of the current components of the CIS. At present, the Agency use of the CIS is limited to two programs, MSSS and MLAB. Both of these are detailed by others at this workshop.

One component of the CIS that is not now being used, but which is of enormous potential use to the Agency, is the chemical substructure search (SSS).* Substructure search allows the scientist to query the file for data of interest to him including mass spectra infrared data, pesticide registration data, toxic substance data, and so forth. His question is specified by a chemical structure, the language of chemistry. The answers are the data whose structures contained the specified substructure.

At present, the Management Information Data Systems Division (MIDSD) is making a pilot version of the SSS option operational for testing and evaluation on mass spectral and pesticides files. In order to do this, MIDSD has obtained a contract with the Chemical Abstracts Service (CAS) in Columbus, Ohio to provide the CAS registry data for these two files. Included in the registry data is a computer readable representation of a chemical structure called a connection table. The SSS program searches this connection table to find the structures that answer the user's query.

In addition to the mass spectral and pesticide files, MIDSD is also obtaining registry data for the oil and hazardous materials file. Other organizations, outside EPA, are also putting the CAS registry data into their files. The ninth edition of the Merck Index, the National Institute of Occupational Safety and Health (NIOSH) list of 25,000 toxic substances and the National Library of Medicine (NLM) Toxline files are examples of data bases that have or will have registry data as part of their files. As more groups within and outside the Agency get registry data, it will be possible to generate and examine interdisciplinary data bases from a structural point of view. One could ask: "For what compounds in the Merck Index are there mass spectra data available?" If registration data and other related administrative data are available with registry data, one can begin to see if certain structures are known pesticides, toxic substances, hazardous materials, and so forth. Structures appearing in different files can be readily located since the search is based on a picture, the chemical structure, not a name. Names, especially trade names, often tell little about the structure of the material. Indeed, the same material in different files can have many different chemical and trade names. One enormous source of data with CAS registry data is the CAS literature files. At present CAS is producing their current five-year subject index with registry data, thus enabling one to search for structural information when searching the literature. Anyone with any experience with chemical nomenclature and the CAS habit of changing its nomenclature should warmly greet such a system.

These examples all deal with existing files. A file of data which is not part of the CIS is infrared (IR) data. A MIDSD sponsored feasibility study of an IR search system, to be published next month, will include the recommendation for the inclusion of CAS registry data as part of any possible future IR system. An IR search system will enable the scientist to have a complementary tool for the identification of organic materials. With the registry data, one will be able to have a link between the IR and mass spectra files, rather than being required to perform separate searches.

The SSS and registry data will allow the Agency to take specific parts of the CIS relevant to Agency needs, build up individual data bases and finally be able to conveniently and inexpensively link them together.

As the pilot SSS project proceeds, the author would be happy to meet with any groups interested in discussing their specific files and needs and in discussing how MIDSD can help coordinate this effort to link the Agency's chemical files.

*See R.J. Feldmann and S.R. Heller, *J. Chem. Doc.*, 12 (1972), 48. Also see R.J. Feldmann, Chapter 3, *Computer Representation and Manipulation of Chemical Information*, ed. M.J. Wipke, S.R. Heller, R. Feldmann and E. Hyde (New York: John Wiley, 1974).

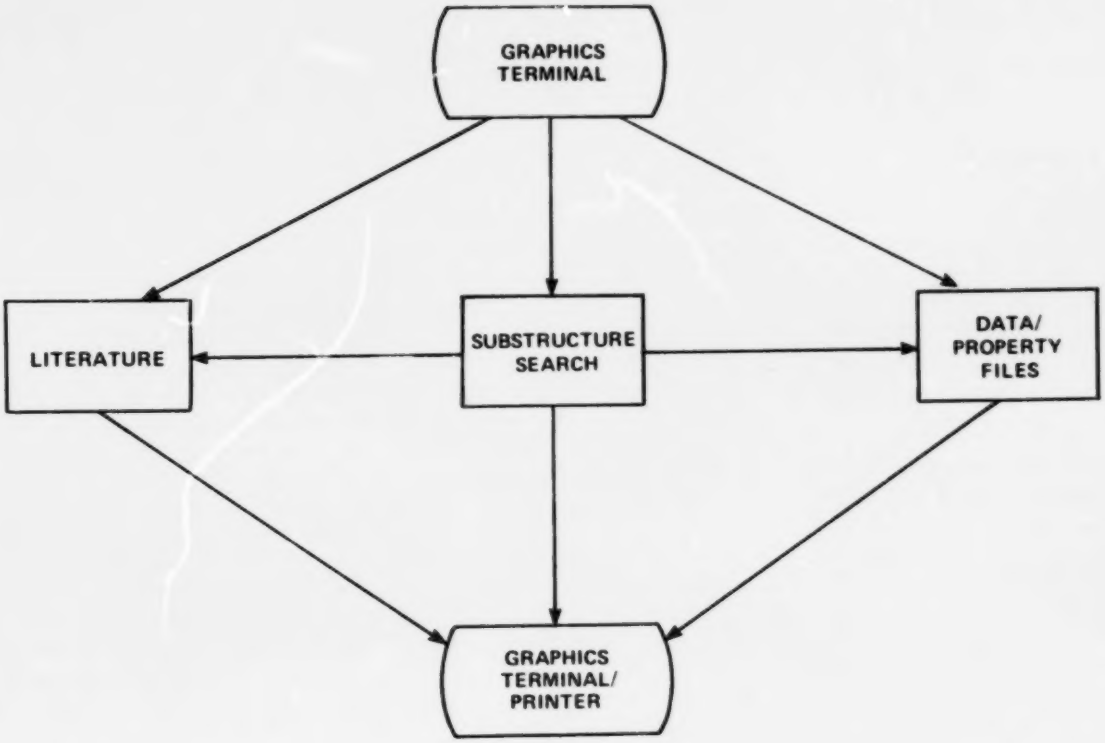


Figure 1
Chemical Information System

```
DCRT/CIS CHEMICAL INFORMATION SYSTEM

TO RUN A PROGRAM IN THE CIS, TYPE THE NUMBER NEXT TO THE NAME OF THE PROGRAM

LITERATURE:

CBAC – 1

STRUCTURE:

SUBSTRUCTURE SEARCH – 2          WLN GENERATION – 3
WLN TO STRUCTURE – 4          WLN GENERATION USING THE TABLET – 17

DATA:

MASS SPEC – 5          NMR – 6

ANALYSIS:

XYZ COORDINATE GENERATOR – 7          XRAY COORDINATE REFORMAT – 8
XRAY MODELING SYSTEM – 9          XRAY CRYSTAL BIBLIO SYSTEM – 10
CNDO/INDO – 11          MINDO – 12          ORTEP – 13
GINA NMR ANALYSIS – 14          ESR SPECTRUM SIMULATION – 15
MLAB CURVE FITTING/MODELING – 16

USER CHOICE:
```

Figure 2
CIS Components

MASS SPECTRAL SEARCH SYSTEMS

By John M. McGuire

INTRODUCTION

Setting and enforcing water quality criteria, determining the fate and effects of water pollutants, and developing optimum control measures require the capability for identifying specific organic pollutants. Table I dramatically illustrates the need to determine the composition of industrial wastes by chemical analysis. The compounds in the left column are those suspected by the discharger, through his knowledge of products, raw materials, and processes, to be in his effluent. The right column, based on chemical analysis of the effluent, contains over twice as many compounds.

Table I
Comparisons of Compounds Reported by Discharger and Compounds Identified by EPA in an Industrial Discharge

Products or Raw Materials Reported	Compounds Identified	
Propylene	allylxylenes*	1,5-cyclooctadiene
Ethylene	isopropylbenzene	styrene*
Butadiene	o-ethyltoluene	o-methylstyrene*
Butane	diacetone alcohol	indan*
Octane	2-butoxyethanol	β-methylstyrene
Ethylene glycol	indene*	dimethylfuran isomer
Ethylene oxide	n-pentadecane	1-methyl indene*
Polyglycols	3-methylindene	acetophenone
Ammonia	n-hexadecane	α-terpineol
Raw gas	naphthalene*	both methyl naphthalene isomers*
Ethane	benzyl alcohol	α-methylbenzyl alcohol
Refinery gases	ethylnaphthalene	phenol*
Refinery C_2 Stream	2,6-dimethyl-naphthalene*	methylethylnaphthalene
Refinery C_3 Stream	cresol isomer	acenaphthene
Propane	acenaphthalene	methylbiphenyl isomer
Hydroformer gas	fluorene	two phthalate diesters
Platformer gas	3,3-diphenyl-propanol	

*Identification was confirmed with a standard.

The identification technique must be highly specific since thousands of compounds of varying degrees of toxicity must be considered. Because some organic compounds are toxic to aquatic organisms at concentrations below 10 μg/ℓ, the technique must also be sensitive.

Gas-liquid chromatography (GC) has adequate sensitivity and reproducibility to provide excellent quantification for volatile organics when the identity of the chemical is known. However, pollutant identifications obtained by comparison of relative retention times are subject to interferences and are questionable for the unknown mixtures found in natural waters. Alternatively, GC may be used as a preliminary separation technique and the effluent may be introduced into a different type of instrument for qualitative identification.

During the 1960's, workers at the Southeast Environmental Research Laboratory (SERL) showed the feasibility of using gas chromatography interfaced with low resolution mass spectrometry (MS) for unknown identification. They used an Hitachi RMU-7 mass spectrometer adjusted for maximum sensitivity, together with manual chart reading and data reduction. Many hours of applied effort are required to gather data, read charts, correct background, construct a data presentation for interpretation (Figure 1), and interpret the data. Because of the large amount of time involved between data collection and interpretation, manual GC/MS is too slow for effective identification of water pollutants.

Most time-limiting factors in manual GC/MS can be accomplished by a computer. To evaluate the feasibility of this approach for EPA laboratories, SERL and the Methods Development and Quality Assurance Research Laboratory in Cincinnati (MDQARL) purchased identical computerized GC/MS systems in 1971. Since that time, more than 20 laboratories have chosen essentially the same system. A PDP-8 minicomputer in this system controls the operation of a quadrapole mass spectrometer and associated output devices.

The applied time required to gather data and prepare it for interpretation was decreased by more than an order of magnitude. The time required for interpretation, however, also had to be shortened.

COMPUTERIZED INTERPRETATION SYSTEMS

Computer-controlled GC/MS produces many spectra from a single environmental sample. Interpretation of these spectra might be done in a variety of ways.

Matrix Inversion

Matrix inversion can be used to give both qualitative and quantitative information if the sample and all

impurities are known to be from a small and well-defined universe. This condition exists in many industrial plants. It does not exist for environmental samples.

Manual Interpretation

Manual interpretation involves assignment of probable compositions to fragments. The assignments are then used to predict the structure of the unfragmented molecule. As a very simple example, Figure 2 shows the assignments in an acetone spectrum. The predominant features in this spectrum are due to mass unit losses of 1(H) and 15(CH_3). Based on the assignments, the molecular structure is strongly suggested. This method is reliable, but it is slow.

Manual Peak Matching

Manual peak matching involves comparing the intensity for each spectral peak with those of literature spectra. Lists such as those of the Mass Spectral Data Centre and of Cornu and Massot give the fastest manual search results; however, use of these lists is nonetheless a slow process. These lists are abbreviated to include only the most intense peaks.

Computerized Spectra Matching

Computerized spectra matching is a practical and rapid way to obtain a probable interpretation. In order to decrease matching times, the spectrum is usually abbreviated by one of several different methods. Biemann and his associates pioneered the 2-most-intense-in-every-14 method illustrated in Figure 3, which shows the Biemann abbreviation for acetone. A comparison of Figures 2 and 3 shows that no significant information has been lost because of the abbreviation.

Computer matching includes many approaches. All require a large library of reference spectra for best performance. The following are some of the most useful:

- The five most intense peaks in the spectrum may be compared with the five most intense peaks in each spectrum of a reference file. The method is very fast, but is not accurate.

- The Biemann abbreviation of the unknown and reference spectra may be compared with intensity information coded to one bit/peak. This program is very rapid on a minicomputer with a disk but can be slow if reference spectra are stored on tape. It is satisfactory for files of less than 600 spectra, but it is not sufficiently selective for larger files.

- The National Institutes of Health (NIH) system developed by Heller, *et al* is an interactive FORTRAN program to perform successive intensity screening based on user-selected peaks. The number of spectra that pass the screen is printed after each screening. Each screening pass is very rapid on the PDP-10. The program includes many user options (Table II). At present, development is being continued jointly by EPA, NIH, and Food and Drug Administration (FDA).

Table II
NIH System Options

TO SEARCH FOR PEAKS, TYPE PEAK
TO SEARCH FOR LOSSES, TYPE LOSS
TO SEARCH FOR MSDC CODES, TYPE CODE
TO SEARCH FOR MOLECULAR WEIGHT, TYPE MW
TO SEARCH FOR MOLECULAR FORMULA, TYPE MF
TO SEARCH FOR PEAKS AND LOSSES, TYPE PL
TO SEARCH FOR PEAKS WITH MSDC CODES, TYPE PC
TO SEARCH FOR PEAKS WITH MW, TYPE PMW
TO SEARCH FOR PEAKS WITH MF, TYPE PMF
TO SEARCH FOR LOSSES AND MSDC CODES, TYPE LC
TO SEARCH FOR LOSSES WITH MW, TYPE LMW
TO SEARCH FOR LOSSES WITH MF, TYPE LMF
TO SEARCH FOR MSDC CODES WITH MW, TYPE CMW
TO SEARCH FOR MSDC CODES WITH MF, TYPE CMF
TO SEARCH FOR MW WITH MF, TYPE MWMF
TO PERFORM A SIMILARITY COMPARISON, TYPE SIM
TO PRINT OUT PEAKS/INTENSITIES AND SOURCE, TYPE SPEC
TO VIEW MICROFICHE, TYPE FICHE
TO PLOT SPECTRA ON DISPLAY TERMINAL, TYPE PLOT
TO COMMENT/COMPLAIN, TYPE CRAB
TO ENTER NEW SPECTRA, TYPE DATA
TO READ THE NEWS OF THE SYSTEM, TYPE NEWS
TO LIST THE MSDC CODES, TYPE LIST
TO EXIT FROM THE PROGRAM, TYPE OUT
USER RESPONSE

- The EPA/Battelle system, which is essentially a faster version of a Massachusetts Institute of Technology (MIT) program, consists of both PDP-8 programs that automatically abbreviate and transmit spectra and also a CDC 6400 spectrum matching program. This assembly language program is very rapid and provides a list of the most probable identifications

ranked by how closely the unknown and reference spectra match. It is essentially an automatic system and provides only a few options.

- A very rapid FORTRAN program for the CDC 6400 has been developed by Clerc and his coworkers in Zurich. It involves automatic screening and ranking of the matches based on a weighted binary-coded signature. Methods for assigning the optimum weighting factors are still being improved.

Learning and Self-Training

Learning and self-training computer approaches offer the best long-range hope of developing a computerized identification program that will not require an extremely large reference library.

The DENDRAL program of the Stanford group is a 360/67 program that does not require a large library. It deduces all possible structures for a given elemental composition and exhaustively tests each structure by predicting a spectrum for it and comparing the predicted spectrum to the unknown one. It is slow.

The Self-Training Interpretive and Retrieval System (STIRS) program is a PDP-11/45 program, developed under NIH and EPA sponsorship, that compares library spectra to the unknown spectrum for various classes of spectra-derived data. The best matches for each class are ranked. This system can lead to identifications even if the reference library does not include a closely related spectrum.

EPA SPONSORED IDENTIFICATION PROGRAMS

Three of the foregoing are of particular interest because of EPA's involvement in their development. These are the NIH system,* the EPA/Battelle system,† and STIRS.††

NIH System

The basic NIH search program, PEAK, permits the user to select an intensity ranging factor and from 0 to 3 mass-intensity pairs (peaks) in each 14-mass interval. After each peak input, the program searches the library mass file for all intensities lying within a window determined by the ranging factor. It then tells the user the number of compounds from the reference library that have passed the screening and gives him the option of obtaining a listing of these compounds. If the number is high, the user can input another peak, which may be from the same 14-mass interval or from a different one. This process is repeated until the user requests a listing. An example of this sequential screening (Figure 4) shows a typical search, which was terminated after the number of references was reduced to less than 20. Five of the six references that passed the screening are isomers of the correct compound.

An important variation of the PEAK program is the LOSS program. In this program, each input is a mass loss and intensity pair. Input losses are calculated by subtracting the mass value from the molecular weight of the compound. Figure 5 illustrates the LOSS search for the same spectrum that was used to illustrate the PEAK search. Two of the ten references are isomers of the correct compound.

The NIH system presently contains three other primary spectra retrieval modes; these are searches for molecular weight, molecular formula, and compound type. The molecular weight search is a screening for all compounds in the file that have a given molecular weight; the molecular formula search can be used to give all compounds having the stated formula, all compounds having stated elements, or all compounds having a stated number of atoms of selected elements. The compound type search involves screening for 86 compound classification codes. All of the screening and retrieval modes can be paired; therefore, it is possible, for example, to perform a LOSS search on a file that consists only of one compound type or a PEAK search on a file of only one molecular weight.

When a search has been narrowed to a reasonably small number of compounds, the NIH program permits ranking the fit of each of these to the input spectrum by means of a similarity comparison. Figure 6 shows dialog and output of this ranking routine for the references found in the PEAK search of Figure 4. The five isomers all have low dissimilarities.

**User's Manual Mass Spectral Search System*, EPA, Washington, D.C. (May 1974).

†*User's Manual Battelle Mass Spectral Matching System*, Battelle Columbus Laboratories, Columbus, Ohio (1974).

††K.S. Kivok, R. Venkataraghavan, and F.W. McLafferty, *J. Amer. Chem. Soc.*, 95, 4185, (1973).

When the number of the desired spectrum has been located in the reference file, the program SPEC can be called to output a listing of the entire spectrum in a typed format.

Two options require certain hardware at the remote terminal. These are the PLOT routine, which permits retrieval and display of any spectrum in the reference library, and the FICHE routine, which locates a desired spectrum on a microfiche copy of the reference library. The PLOT routine for PDP-10 users will permit CALCOMP or ZETA plotted copy, or will display on any of six graphics terminals. The FICHE routine either provides information for manual location or drives a COMCARD viewer to display the desired spectrum.

Other present NIH options permit the user to enter data, obtain system news, and comment to the system manager.

EPA/Battelle Matching Program

The EPA/Battelle matching program, taking advantage of the high information redundancy of mass spectra, is based on the two most intense peaks in every 14 mass units. There are four main steps in the matching process:

- Screening based on molecular weight range
- Screening based on the most intense peak of the unknown spectrum
- Presearching based on the spectrum family
- Ordering the best matches based on peak-by-peak comparison of the unknown spectrum with those reference spectra passing the presearch.

To reduce operation time and eliminate human errors and prejudices in selecting, formatting, and transmitting data, PDP-8 utility routines transfer input spectra data directly from the user's remote PDP-8 to the central CDC 6400. These programs have been evaluated and improved during the past year. A match against the present data base of 9000 spectra (8400 general organic spectra culled from the Aldermaston collection and 600 pollutant spectra from SERL and Battelle) requires approximately 40 seconds of elapsed time.

The similarity index (SI) gives the user an immediate indication of the quality of the matches. The "best hit" will be the first identification; the SI will show whether it is a poor match (<0.2 if the data base does not contain any closely related compounds), one of several fair matches (0.2-0.35 if the correct compound is not in the data base but related ones are), or a good match (>0.35 if the SI of the second best hit is significantly lower).

In one study made at SERL, 50 percent of the unknowns present in the effluent of a kraft paper mill were found correctly as the best hit, 8 percent as the second best hit, and 2 percent as the third best hit (13). The success of the system should improve since reference spectra are added continually.

The search may miss the match because of poor quality data in the input spectrum. If the spectrum looks poor, or a match is not found using the standard search parameters, the user has the option of decreasing different portions of the presearch selectivity. This is done by redefining parameters for ratio and rectangular array in the search dialog.

Compared with magnetic deflection spectrometers, quadrupole instruments exhibit a bias toward low mass. This is demonstrated in Figure 7, which compares both types of spectra for the pesticide, parathion. Since the Aldermaston data base is comprised primarily of spectra obtained on magnetic deflection mass spectrometers and most EPA laboratories have quadrupole instruments, a major concern in the development of the matching system was whether suitable matches could be obtained between quadrupole and magnetic deflection spectra. Experience with the system has shown that the program provides excellent matches for both quadrupole and magnetic deflection spectra.

The same magnetic deflection spectrum of p-ethyltoluene that was used to illustrate the NIH program (Figure 4 through Figure 6) can be used as well as an example of the EPA/Battelle program. Figure 8 shows the results for this search. All of the best hits are isomers; however, the first and third matches are most closely related to the correct structure.

Another example of the use of the EPA/Battelle matching program is provided by a study of the effluent of an experimental coal gasification plant. Organic components were extracted with methylene chloride. The

chromatogram of the extract (Figure 9) contained seven distinct peaks.

In a computerized matching of the quadrupole spectra for those compounds, the best matches were with C_6, C_7, and C_8 hydroxyl-containing materials. The output indicated high SI's for the first six peaks, but a low one for the last GC peak. Subsequent visual inspection of the mass spectrum for this GC peak indicated that the last peak arose from two compounds with the same retention time.

The identifications are given in Table III. When different materials were selected by the matching program as the best and second best matches, relative GC retention times favored the best match over the second best. In a continuation of the computer dialogue, given in Figure 10 for GC peak 3, thirteen cresol spectra were matched with SI's greater than 0.645. The first non-cresol match was 3-tolyl-N-methyl carbamate with an SI of 0.574.

Table III
Compounds Tentatively Identified in Waste Effluent of Coal Gasification Pilot Plant

GC Peak	Best Match	SI	Second Best Match	SI
1	Phenol	0.834	Phenol	0.795
2	o-Cresol	0.846	m-Cresol	0.741
3	m-Cresol	0.867	o-Cresol	0.842
4	2,5-Dimethylphenol	0.839	2,6-Dimethylphenol	0.804
5	3,4-Dimethylphenol	0.700	3,4-Dimethylphenol	0.692
6	2,4-Dimethylphenol	0.653	3,4-Dimethylphenol	0.637
7	α-Naphthol	0.245	1,2-Dihydroxynaphthalene 1,2-Dihydronaphthalene	0.232

Table IV summarizes results obtained by Budde and Eichelberger* of the MDQARL in an objective evaluation of the NIH and EPA/Battelle systems, using a mixture of typical environmental pollutants.

STIRS

Because of the relatively small size of both the NIH and EPA reference libraries, a substantial number of environmental samples cannot be matched by either of these systems. The STIRS deduces substructures based on reference spectra and permits the analyst to reconstruct the structure of the unknown even if the library does not contain a matching spectrum. The program permits the user to transmit a complete mass spectrum from a remote PDP-8 to the Cornell PDP-11/45 via the same serial data interface and acousticoupler used by the EPA/Battelle system. The input peaks are compared to spectra in the 25,000 spectra reference library and used to calculate the probable molecular weight and value (from 0 to 1000) for 14 match factors. These factors are based partly on known mass fragmentation behavior and partly on experience gained through the use of this program. For each of these match factors (Table V), the 10 best matches in the data base are ordered by spectral similarity and listed for the user. A 15th match factor is also generated as the overall match factor.

A search using STIRS gives not only the chemical name, but also Wiswesser line notations (WLN's) for each of the compounds selected for the particular match factor. Figure 11 is the STIRS output obtained by transmitting a spectrum of ethyltoluene. The confidence table indicates a high confidence that a phenyl ring is present, a lesser confidence that an ethyl group is present, and a low confidence that a methyl group is present. All of these conclusions are correct. Low confidence suggestions of unsaturation and a branched chain are incorrect. A set of criteria is being developed that will permit predicting the probabilities that given groupings are present in the unknown; however, scanning the WLN's found for each factor shows that certain WLN substructures are repeated far more frequently than others. If the same substructure is indicated by several match factors and another substructure by other match factors, the correct structure probably contains both substructures. This intuitive conclusion has been demonstrated to be correct in many cases.

STIRS suffers to a lesser degree from the same library limitation as do the EPA/Battelle and NIH systems. A spectrum of the exact material or even a closely related one does not have to be present in the reference library, but some compounds having the appropriate substructures must be present. A spectrum for a phosphorthioate was found to give no useful information when it was first submitted to STIRS. This led to the discovery that none of the spectra in the library at that time included the phosphorthioate substructure.

Recently, an American Society for Mass Spectrometry panel submitted identical test spectra to the Mass Spectral Data Centre, NIH, EPA/Battelle, and STIRS searches. The results showed STIRS to be markedly superior to all of the others in its ability to suggest correct identifications.

*W.L. Budde and J.W. Eichelberger, personal communication. July, 1974.

Table IV
Summary of Results Obtained with Some Typical Organic Water Pollutants Using the NIH and EPA Mass Spectrum Matching Programs

Typical Water Pollutant	NIH (Operator Selected Ions)			EPA (Automatic Abbreviated Spectrum)		
	No. of Mass/Abundance Pairs Entered[a]	No. of File Spectra Matched (Hits)	No. of Correct Hits[b]	No. of File Spectra Matched (Hits)	Rank of the Correct Hits by Similarity Index[b]	SI
1,1,2,2-tetrachloroethane	3	7	3	13	1	.567
Phenol	5	8	4	24	1,2	.878,.691
bis (2-chloroisopropyl)ether	6	1	1	65	1,2,3	.758,.758,.602
p-cresol	11	10	2[c]	78	1	.808
α-terpineol	5	4	2	0	[h]	
naphthalene	5	7	3	18	2,3	.771,.758
o-nitrotoluene	3	6	0[d]	15	1,2	.892,.586
benzothiazole	4	2	1	25	1,2	.732,.728
2-methylnapthalene	10	11	5[e]	8	2[g]	.668
1-methylnaphthalene	3	3	2[f]	9	1,2	.609,.598
hexadecane	5	8	3	442	3	.766
acenaphthene	3	8	5	17	1,2	.786,.753
dibenzofuran	4	8	3	11	1,2	.682,.560

[a]No more than two mass/abundance pairs were entered for each standard NIH 14 amu interval. The data from the highest mass interval observed was entered first followed by the next highest, etc. Within each interval the higher mass number was always entered first. Input was stopped when 10 hits or less were reported or the data was exhausted. With this system a different sequence of data entry can give significantly different results.

[b]Duplicate correct answers indicate there are duplicate spectra in the data base.

[c]Seven other hits were either m-cresol or o-cresol which have very similar mass spectra.

[d]The spectrum of o-nitrotoluene was in the data base but was in error.

[e]The other six hits were 1-methylnaphthalene which has a virtually identical mass spectrum.

[f]The other hit was 2-methylnaphthalene which has a virtually identical mass spectrum.

[g]The hit with the highest SI (.743) was 1-methylnaphthalene which has a virtually identical mass spectrum.

[h]Remote computer aborted three times with three different abbreviated spectra from two different GC/MS runs. Manual input of one of the same data sets several days later yielded four hits including two correct with an SI of 1.00.

Table V
Spectral Data Classes and Match Factors Utilized by STIRS*

	Class of Spectral Data	Match Factor
1.	Ion series	MF 1
2.	Low mass characteristic ions (< m/e 90)	MF 2
2B.	Overlapping characteristic ions[a] (m/e 67-115)	MF 12
3.	Medium mass characteristic ions (m/e 90-149)	MF 3
3B.	Overlapping characteristic ions[a] (m/e 119-175)	MF 13
4.	High mass characteristic ions (> m/e 150)	MF 4
5.	Neutral loss series[1]	MF 14
6.	Small primary neutral losses	MF 5
6B.	Overlapping primary neutral losses[a] (33-90)	MF 15
7.	Large primary neutral losses	MF 6
8.	Secondary neutral losses from the most abundant odd mass loss	MF 7
9.	Secondary neutral losses from the most abundant even mass loss	MF 8
	Class 9 data of the unknown spectrum matched against Class 5 data of the reference spectrum	MF 9
10.	Fingerprint ions	MF 10
	Overall match factor	MF 11

[a]Proposed new match factor currently under experimental evaluation.

[1]A. McLafferty, "Computer-Aided Interpretation of Mass Spectra (I)," in *Gas Chromatography/Mass Spectrometry in the Investigation of Human Disease*, ed. O.A. Mamer and C.R. Scriver (Montreal, Canada: McGill University-Montreal Children's Hospital, 1974).

FUTURE PLANS

Work underway and plans for the future include joint efforts by the Office of Research and Development (OR&D), Management Information and Data Systems Division, NIH, and FDA to:

- Expand and centralize a large, refined data base
- Translate the EPA/Battelle program to FORTRAN for inclusion with the NIH programs
- Develop software and hardware to permit automatic data transmission between all EPA GC/MS/computers and the NIH PDP-10
- Make all matching programs developed available to the public through cooperation of the Mass Spectral Data Centre
- Expand the capabilities of the EPA/Battelle system to maintain geographic frequency distribution of all identified organic pollutants
- Develop FORTRAN procedures for deconvolution of overlapping GC/MS peaks
- Expand STIRS to allow the program to predict the most probable structure.

All of these search programs are useful to EPA in determining the identity of pollutants. This utility will be greatly increased as the above plans are completed. STIRS has the greatest potential of any of the EPA programs to provide rapid identification of any GC/MS unknown with a library of less than 50,000 spectra. As is true, however, for all computer identifications, such identifications are suggestions of possible structures and do not remove analytical responsibility from the chemist. Final identification requires confirmation by the analyst.

Figure 1
Acetone Spectrum

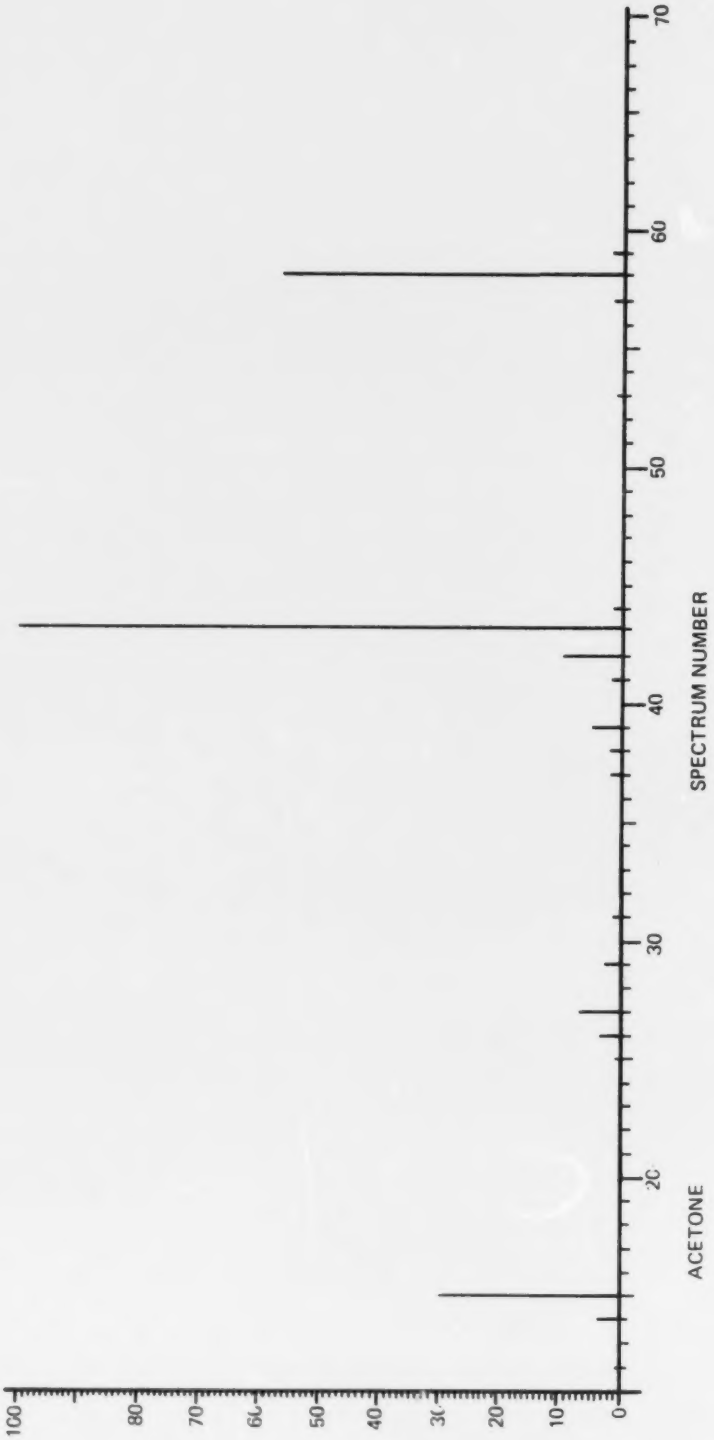


Figure 2
Acetone Spectrum - Major Peaks Assigned

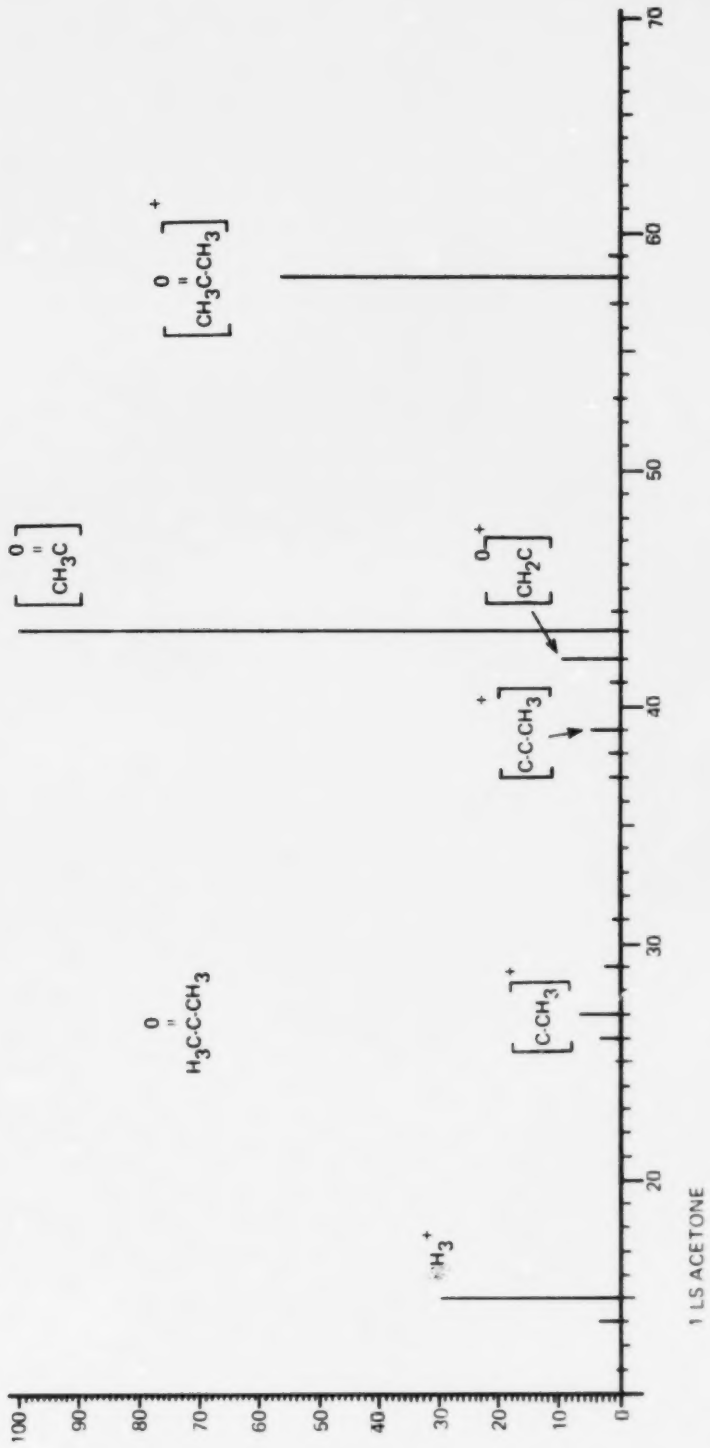
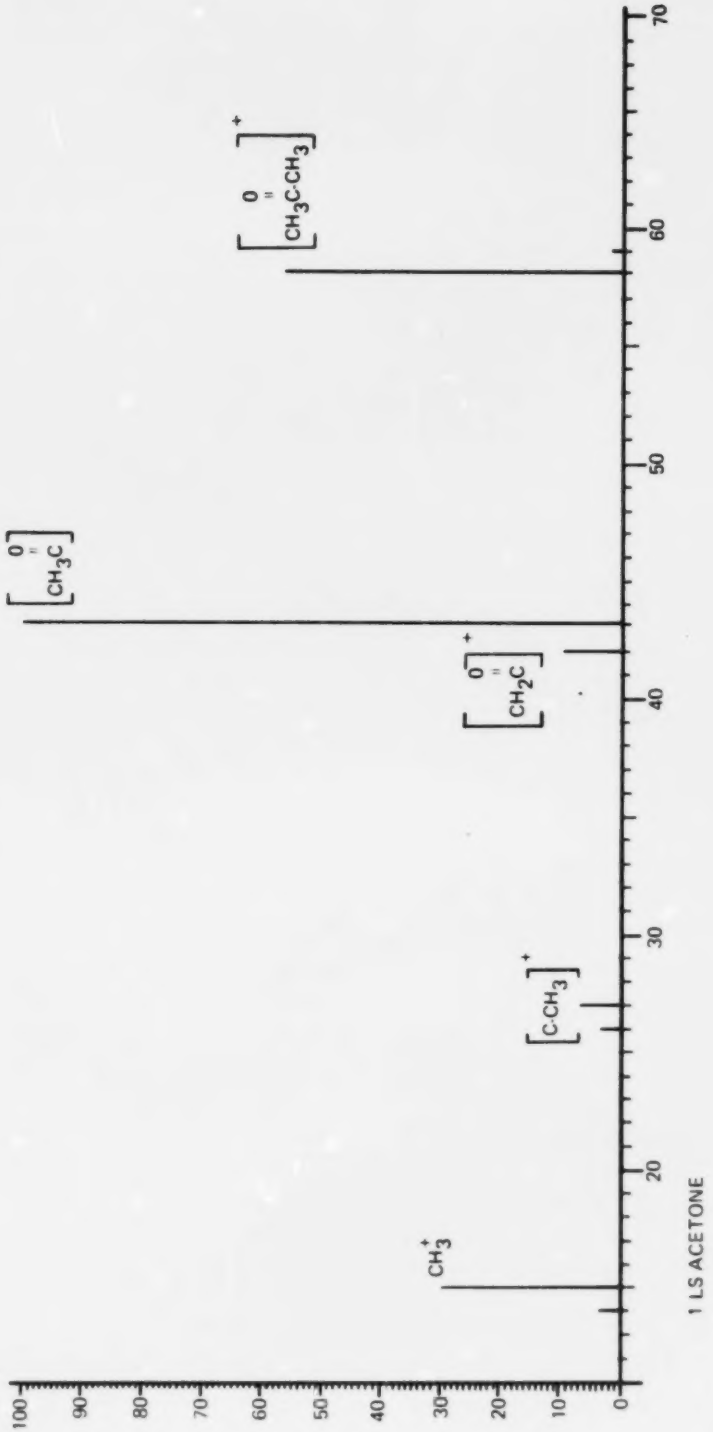


Figure 3
Acetone Abbreviated Spectrum



```
TYPE PEAK, INT
CR TO EXIT, 1 FOR ID,MW,MF, AND NAME

USER 1Ø5,1ØØ

   # REFS      M/E PEAKS

     893          1Ø5

NEXT REQUEST 12Ø,38

   # REFS      M/E PEAKS

      96          1Ø5 12Ø

NEXT REQUEST 91,11

   # REFS      M/E PEAKS

      25          1Ø5 12Ø  91

NEXT REQUEST 79,16

   # REFS      M/E PEAKS

       7          1Ø5 12Ø  91  79

NEXT REQUEST  1



  ID#      MW       MF                NAME
 2135      12Ø C9 H12              ISOPROPYLBENZENE
 4867      152 C9 H12 02           CUMENE HYDROPEROXIDE
16799      388 C12 H12 03 W        PI-1,3,5-TRIMETHYLBENZENE-
                                   TRICARBONYL TUNGSTEN
2Ø472      12Ø C9 H12              ISOPROPYLBENZENE
2Ø473      12Ø C9 H12              ISOPROPYLBENZENE
2Ø474      12Ø C9 H12              1-METHYL-2-ETHYLBENZENE

2Ø475      12Ø C9 H12              O-ETHYLTOLUENE

NEXT REQUEST
```

Figure 4
NIH PEAK Search for p-Ethyltoluene

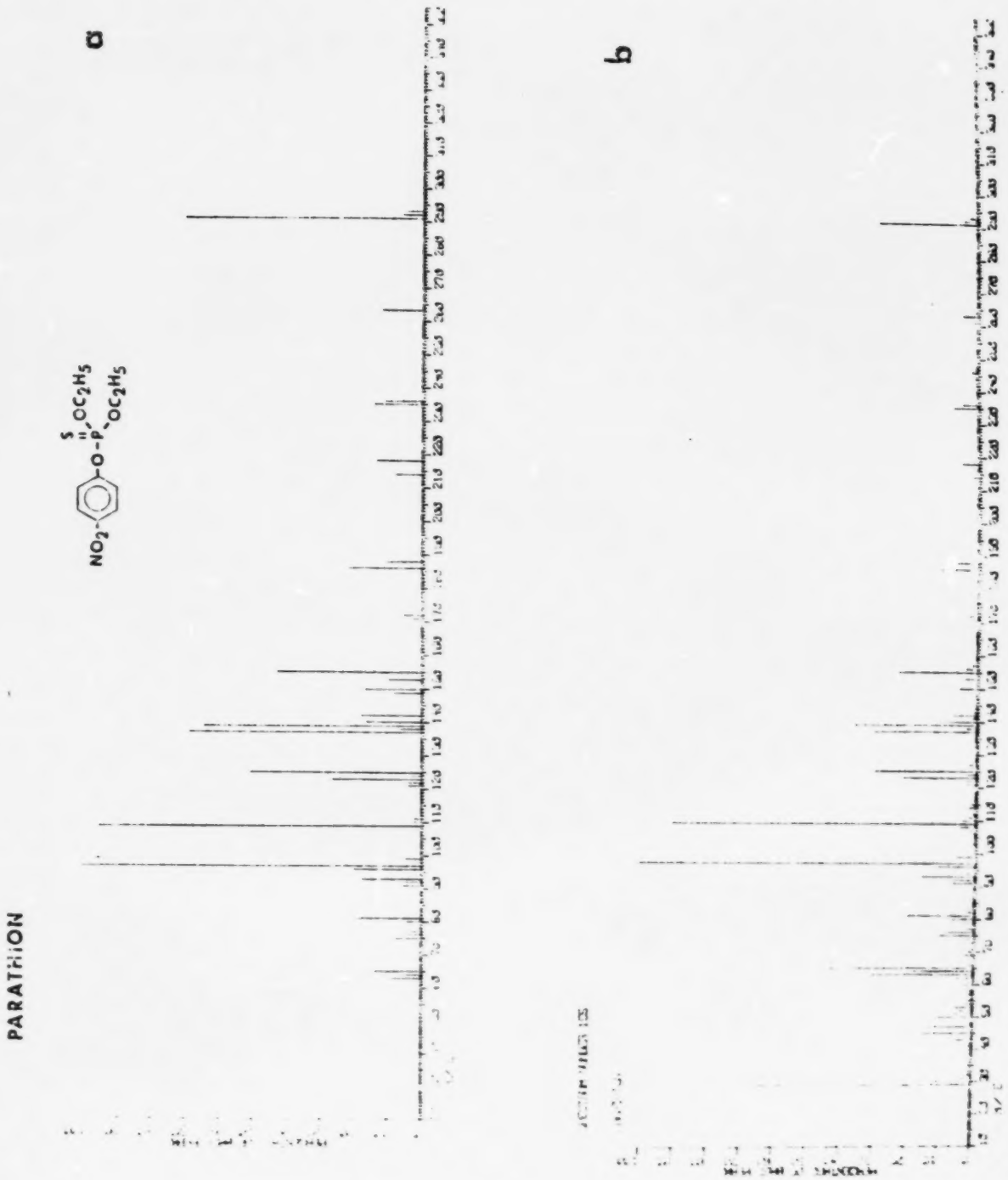
```
TYPE LOSS, INT
1 TO EXIT, 2 FOR ID,MW,MF AND NAME

USER -15,1ØØ

  # REFS        LOSSES

   2169        -15

NEXT REQUEST:  -29,12

  # REFS        LOSSES

    5Ø4        -15 -29

NEXT REQUEST:  -41,16
  # REFS        LOSSES

     84        -15 -29 -41

NEXT REQUEST:  -43,18

  # REFS        LOSSES

     39        -15 -29 -41 -43

NEXT REQUEST:  -55,7

  # REFS        LOSSES

     1Ø        -15 -29 -41 -43 -55

NEXT REQUEST:  2

  ID#      MW      MF                NAME

  241       7Ø C5 H1Ø              3-METHYL-1-BUTENE
  26Ø       7Ø C4 H6 0             3-BUTYN-2-OL
  446       82 C6  H1Ø             1-METHYLCYCLOPENTENE
 5657      16Ø C12 H16             1-(2,3-DIMETHYLPEHNYL)-BUT
                                   -2-ENE
19Ø76       7Ø C5 H1Ø              2-METHYL-2-BUTENE
19271       82 C6 H1Ø              METHYLENE CYCLOPENTANE

19272       82 C6 H1Ø              METHYLENE CYCLOPENTANE

19274       82 C6 H1Ø              CYCLOHEXENE
20476      12Ø C9 H12              M-ETHYLTOLUENE
TYPE C TO CONTINUE LIST,
CR TO EXIT:  C

2Ø477      12Ø C9 H12              1-METHYL-3-ETHYLBENZENE
```

Figure 5
NIH Loss Search for p-Ethyltoluene

STD ETHYL ID#	DISSIMILARITY
2135	Ø.14
4867	25.Ø1
16739	6.83
2Ø472	Ø.12
2Ø473	Ø.11
2Ø474	Ø.1Ø
2Ø475	Ø.16
USER RESPONSE	

Figure 6
Dissimilarity Comparision for PEAK Search of Fig. 4

Figure 7
Comparison of Parathion Spectrum From a) Magnetic Instrument and b) Quadrupole Instrument

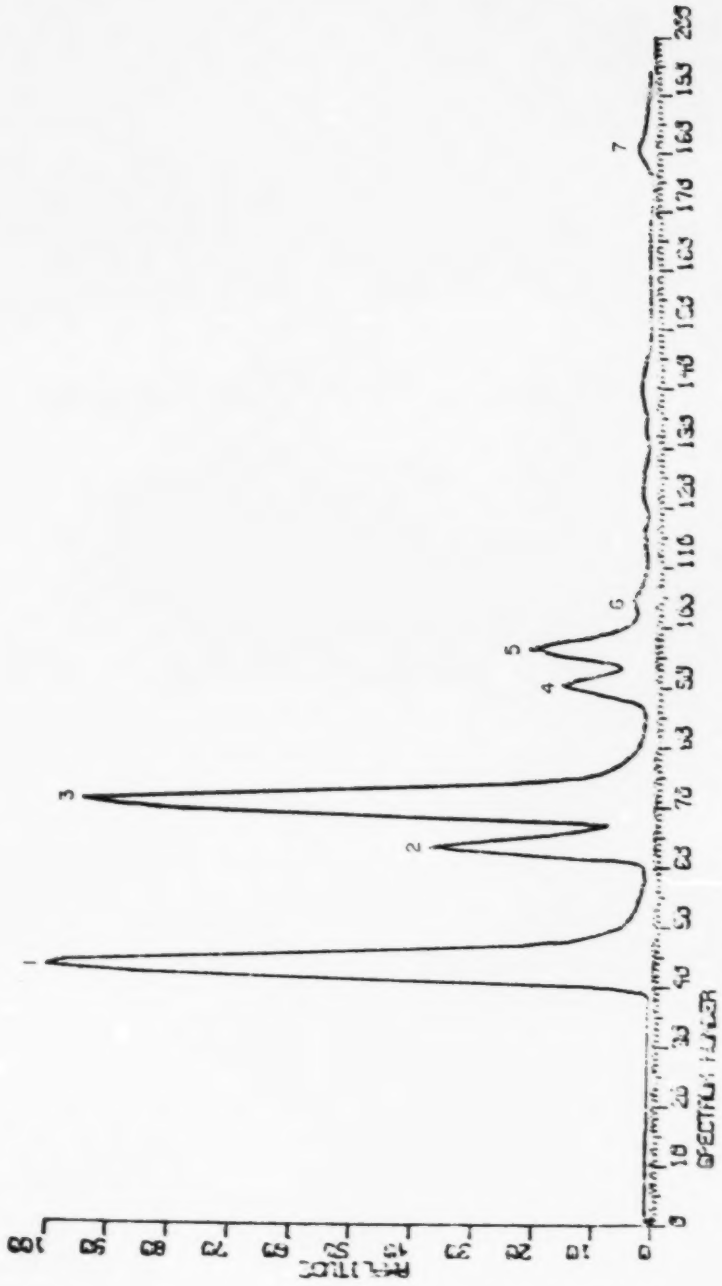


Reproduced from best available copy.

```
I.D.?  ETHYLTOLUENE
PAPER TAPE?      N
MANUAL
STANDARD
27,4;39,8;41,3;51,8;58,4;63,4;65,7;77,18;^
79,16;91,12;1Ø3,1Ø;1Ø5,1ØØ;1Ø6,9;12Ø,38;121,4;^
END
PARMTRS?         M1ØØ-2ØØ
 74 HITS
1-METHYL-2-ETHYLBENZENE 12Ø C9 H12 API Ø312
FILE KEY= 3136
SI=Ø.7ØØ

ISOPROPYLBENZENE (CUMENE)  12Ø C9 H12 API Ø311
FILE KEY= 3135
SI=Ø.691

1-METHYL-3-ETHYLBENZENE 120 C9 H12 API 0313
FILE KEY= 3137
SI=Ø.66Ø

1,2,3-TRIMETHYLBENZENE 12Ø C9 H12 API Ø315
FILE KEY= 3139
SI=Ø.631

1,2,4-TRIMETHYLBENZENE 12Ø C9 H12 API Ø316
FILE KEY= 314Ø
SI=Ø.569
```

Figure 8
EPA/Battelle Matching of p-Ethyltoluene

Figure 9
Reconstructed Gas Chromatogram of Coal Gasification Plant Extract



```
S, E, ØR P?S
I.D.?  GØAL GASIFICATION PLANT EFFLUENT
PAPER TAPE?Y
FN--F73  ; S   -1  :
CHPFI-1 (1ST EXT)  :
37,3;38,8;39,39;40,4;41,2;43,2;50,12;51,20;52,10;61;:
53,18;54,6;55,4;61,2;62,4;63,9;64,2;65,4;66,2;56;:
74,2;77,41;78,9;79,38;80,13;81,2;89,4;90,12;91,6;61;:
106,3;107,100;108,91;109,6;33;:
END

PARMTRS?  M100-500
111 HITS
M-CRESØL 108 C7.H8.Ø  AST 0181
FILE KEY= 186
SI=0.857

1-HYDRØXY-3-METHYLBENZENE (3-METHYLPHENØL--M-CRESØL)
108 C7.H8.Ø  TRC 0068
FILE KEY= 6392
SI=0.845

1-HYDRØXY-2-METHYLBENZENE (2-METHYLPHENØL--O-CRESØL)
108 C7.H8.Ø  TRC  0067
FILE KEY= 6391
SI=0.834

1-HYDRØXY-4-METHYLBENZENE (4-METHYLPHENØL--P-CRESØL)
108 C7.H8.Ø  TRC 0069
FILE KEY= 6393
SI=0.815

M-CRESØL 108 C7.H8.Ø  AST 0459
FILE KEY= 462
SI=0.805
```

Figure 10
Computerized Spectra Matching Program Dialogue for a Component of Coal Gasification Plant Extract

```
HIGH MASS FINGERPRINT ION MATCH   (MF1Ø)

  MF1    MF2    MF3    MF4    MF5    MF6    MF7    MF8    MF9   MF1Ø   MF11

  711    4Ø9    622     Ø     766     Ø     833    75Ø    25Ø    675    678
# 2139     1-METHYL-2-ETHYLBENZENE
2R B     C9 H12       12Ø

  711    4Ø9    632     Ø     692     Ø     833    75Ø    25Ø    637    644
# 2141     1-METHYL-4-ETHYLBENZENE
2R D     C9 H12       12Ø

  711    5Ø7    649     Ø     857     Ø     833    75Ø    25Ø    635    743
# 2135     ISOPROPYLBENZENE
1YR      C9 H12       12Ø

  533    23Ø    8Ø5     Ø     23Ø     Ø     666     Ø      Ø     634    411
# 5848     2-PHENYLHEXANE
4YR      C12 H18      162

  666    347    715     Ø      83     Ø     5ØØ     Ø      Ø     624    346
# 5185     ALPHA-CHLOROACETOPHENONE
G1VR     C8 H7 O1 CL1       154

OVERALL MATCH   (MF11)

  MF1    MF2    MF3    MF4    MF5    MF6    MF7    MF8    MF9   MF1Ø   MF11

  711    5Ø7    649     Ø     857     Ø     833    75Ø    25Ø    635    743
# 2135     ISOPROPYLBENZENE
1YR      C9 H12       12Ø

  711    4Ø9    622     Ø     766     Ø     833    75Ø    25Ø    675    678

# 2139     1-METHYL-2-ETHYLBENZENE
2R B     C9 H12       12Ø

  711    4Ø9    617     Ø     766     Ø     833     Ø      Ø     617    677
# 214Ø     1-METHYL-3-ETHYLBENZENE
2R C     C9 H12       12Ø

  711    579    496     Ø     764     Ø    1ØØØ   1ØØØ    25Ø    589    677
# 2133     6-METHYL-6-ETHYL FULVENE
L5YJ AUY2        C9 H12       12Ø

  711    4Ø9    632     Ø     692     Ø     833    75Ø    25Ø    637    644
# 2141     1-METHYL-4-ETHYLBENZENE
2R D     C9 H12       12Ø

                 % CONFIDENCE TABLE -- BASED ON WLN'S
SEARCH LIMITS  9Ø-1ØØ
SYMBOL MATCH FACTOR  CONFIDENCE

CH3   MF 7  99     MF11   94

ETH   MF 5  99     MF 7   99     MF11    99

PHN   MF 1 1ØØ     MF 2  1ØØ     MF 3   1ØØ    MF1Ø   1ØØ    MF11   1ØØ

C=C   MF 7 1ØØ

Y     MF 7  99     MF 8   99
```

Figure 11
Typical STIRS Output

BOWNE TIME SHARING (WORD/ONE)

By Catherine Tittle

Large volume, impossible deadlines, small staffs and limited budgets traditionally create a significant roadblock in the preparation of research reports, manuals, directories and other documents. The Environmental Protection Agency (EPA), faced with this problem since it was organized, found the solution in Word/One.

The EPA has been using Word/One since 1971. Beginning in Washington and spreading to the Regions and National Environmental Research Centers, Word/One has been used successfully by EPA to produce many of the regulations, manuals, reports, directories, mailing lists and so forth which comprise a significant portion of the Agency's paperwork production.

Five years ago, Bowne Time Sharing (BTS) began to apply computer technology to the business of producing words rather than crunching numbers. Word/One, BTS' computerized text-processing system, was designed to deal with the ever increasing paperwork explosion. (Certain applications lend themselves more readily and cost-effectively to computerized word processing: Figure 1.) Using the shared computer approach, Word/One permits the simultaneous use of a central computer by many remote locations using a typewriter/terminal and a telephone line. This service is available wherever phone service exists and is supported by Bowne Time Sharing's seven offices in Washington, D.C., Boston, New York, Chicago, Philadelphia, Atlanta and Los Angeles.

The BTS computer configuration, which is located in New York, is an IBM 370, model 155, with a million bytes of core storage. The peripheral devices, which are used for on-line storage, are capable of providing immediate access to over a billion bytes of text information. The printing volume, which is naturally the highest volume of output, is handled by five high-speed printers that operate at a speed of over 400 lines per minute each and produce 800,000 to 1 million lines of typewriter quality, upper and lower case, information daily. Additional output can include magnetic tape and punched cards. Figure 2 shows the interrelation of the different parts of BTS' Word/One System configuration.

Word/One is designed for use by personnel having little or no computer knowledge or background. All instructions are phonetic.

The system consists of four basic segments:

- Input
- Storage and Retrieval
- Editing and Manipulation
- Format and Print.

A discussion of each segment follows.

Input

Word/One is designed to: (1) capture keystrokes, thereby reducing typing of future drafts on magnetic disk and (2) eliminate (mechanical and error-prone) typing tasks, including insertion of heading information and page numbering.

Storage and Retrieval

Word/One allows the user to: (1) store, for an indefinite time, any input material, (2) assign a convenient name, (3) retrieve, immediately, any of the material, (4) optionally share the document with selected other organizations for joint efforts and (5) maintain document records, e.g. the date data is stored, for complete management control.

Editing and Manipulation

Word/One is designed to: (1) minimize typing when revising a document, (2) eliminate redundant proofreading of previously approved material, and (3) reduce the delay in producing the next draft. Selected Word/One features that assist in this area are:

- Replace: Word/One can change all occurrences of a word or phrase to another at all or preselected locations.
- Elasticity: Word/One will automatically expand or contract each edited segment of a document to accommodate the change.
- Proofmark: Word/One, for working drafts, automatically places a special character to the right of each edited line so that the reader will know where each change occurred.

Format and Print

Word/One provides a number of text formatting options that allow: (1) mechanical functions, such as insertion of headings to be specified rather than typed each time, (2) error-prone operations, such as the maintenance of tabular material on a page, to be handled by the system, (3) flexibility in changing the document layout, such as page width and depth specifications, and (4) print output options, such as justified left and right margins.

In addition, the Word/One system provides a significant level of throughput. The high-speed printers produce typewriter equivalent copy overnight, regardless of document size. Most of the low-speed terminals currently used by EPA are compatible with the Word/One system.

Using this concept EPA has produced many publications including EXPRO, the OR&D Program Planning and Reporting Manual, Methods Manual for Chemical Analysis of Water Wastes (1974), and development documents and regulations in several program areas. These publications were produced meeting tight deadlines with limited clerical staffs. Productive time of both clerical and professional personnel involved was maximized because the material was typed only once, changes were made only where necessary and the high-speed printer was used to speed turnaround.

Where large documents with significant revision are handled, Word/One has proven to be a necessary and beneficial tool.

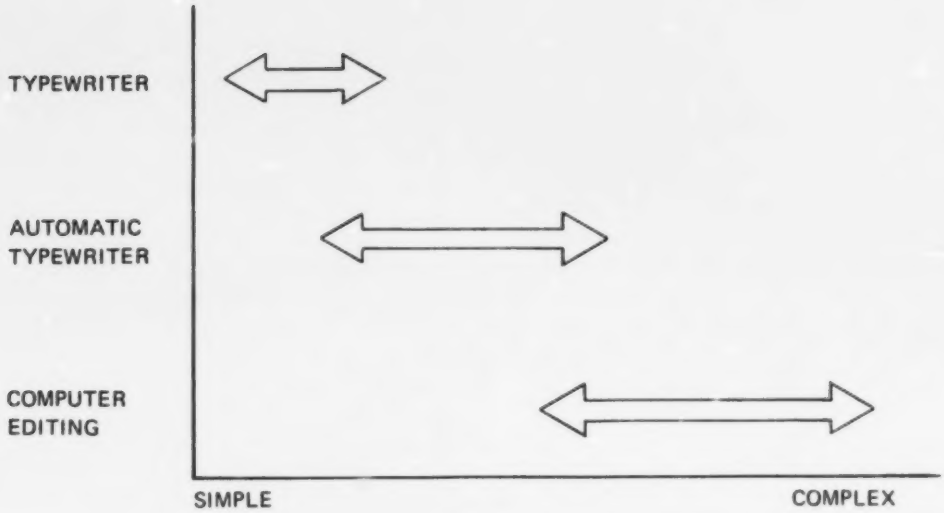


Figure 1
Equipment Method Cost-Effectiveness Range vs Text Manipulation Requirements

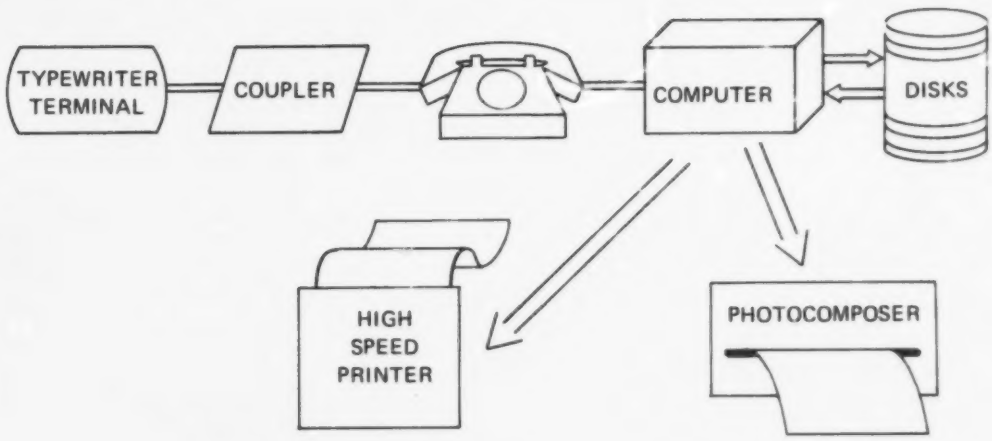


Figure 2
Equipment and Data Flow for Word/One Applications

INTRODUCTION TO THE UNIVAC 1110 AT RESEARCH TRIANGLE PARK - CAPABILITIES

By T. L. Rogers

INTRODUCTION

In June 1973 UNIVAC was awarded the contract to provide the new computer system for the Research Triangle Computing Center. The new system as installed (Figure 1) had to pass a benchmark test with a throughput ratio of 12 to 1 as compared to the IBM 360/50. The system is a UNIVAC 1110 with the components and general capabilities as described in this paper.

GENERAL DESCRIPTION

The UNIVAC 1110 System is a general purpose, high performance system incorporating the latest advances in computer design, system organization, and programming technology. The various components of the UNIVAC 1110 System are designed as separate logical units providing maximum functional modularity. The multiprocessing capabilities are an integral part of the system; the command/arithmetic units can perform numerous tasks simultaneously under control of a common executive. The flexible modular structure enables a user to tailor a system to his individual requirements. Principle features of the UNIVAC 1110 System are:

- Common resource systems organization
- Multiple command/arithmetic units (CAUs) and input/output access units (IOAUs)
- Character manipulating instructions
- Partial-word, double-word, and full-word addressability
- System partitioning capability
- Redundancy among system components
- Two levels of directly addressable storage
- Large modular plated-wire primary storage
- Large modular core extended storage
- Storage protection
- Program address relocation
- Independent input/output access units (IOAUs)
- Extensive software library and language processors
- Dynamic adjustment to a mix of batch, demand, and real-time modes
- Wide choice of high performance peripheral subsystems
- Independent, simultaneous communications processing.

SYSTEM COMPONENTS

Each component in the UNIVAC 1110 System is functionally independent and may have the following properties:

- Two or more access paths
- Access conflicts resolved by priority logic.

In a multiprocessor configuration the following capabilities are standard:

- Continued system operation if any component fails
- Any component can be logically removed for servicing without disabling the entire system.

The UNIVAC 1110 System consists of eight types of components:

- Command/arithmetic units
- Input/output access units
- System console
- System partitioning unit

- Primary storage
- Extended storage
- Maintenance controller
- Peripheral subsystems.

Command/Arithmetic Unit (CAU)

The basic UNIVAC 1110 System configuration includes one command/arithmetic unit (CAU) and one input/output access unit (IOAU). All control and arithmetic functions are executed by the CAU. The CAU is a multitask instruction-stacking device capable of controlling up to four instructions at various stages of execution. A CAU can interface with up to four primary storage units by means of both an instruction path and an operand path. Dual data paths connect the CAU with extended storage through a maximum of eight UNIVAC multiple access interface (MAI) units. The data paths to primary and extended storage have overlapping and interleaving capabilities. In a multiprocessor system, the user can specify and can change which units are to be used, thereby permitting a system to be logically divided into two or three independent smaller systems, or removing individual units for maintenance without affecting the total system. Interrupt signals may be sent or received on one of the three interprocessor lines.

Additional features of each CAU are:

- Capability of executing up to 1.8 million instructions per second
- 300-nanosecond effective basic instruction time
- Four-deep instruction stack
- 112-word general register stack (GRS)
- Character manipulation by means of byte-oriented instructions.

The Research Triangle Park (RTP) System has two CAUs.

Input/Output Access Unit (IOAU)

The basic UNIVAC 1110 System configuration includes one IOAU. The IOAU controls all transfers of data between the peripheral devices and primary and extended storage. Transfers are initiated by a CAU under program control. The IOAU includes two nonconcurrent data transfer paths, one for primary storage and one for extended storage.

The IOAU consists of two sections: a control section and a section containing from 8 to 24 input/output channels. Input/output (I/O) data transfers may occur simultaneously with the execution of programs in the CAU.

The control section includes all logic associated with the transfer of function, data, and status words between primary or extended storage and the subsystems. It also services I/O requests from either one or both of the CAUs (in a multiprocessor system) and routes interrupts to one of the two CAUs. Interrupt routing may be specified by program.

Some outstanding features of the IOAU are:

- Aggregate transfer rate of 4 million 36-bit words per second (24 million characters)
- Externally specified index (ESI) and internally specified index (ISI) transfer modes on any channel
- Data chaining
- Interrupt tabling
- Storage-to-storage transfers.

The RTP System has one IOAU.

System Console

The system console provides the means for communication with the executive system. The basic console consists of the following major components:

- The cathode ray tube (CRT)/keyboard consists of a UNISCOPE 100 display terminal. The display format is 16 lines with 64 characters per line. The seven-bit ASCII character set, consisting of 95 characters plus the space, is used. The keyboard provides all of the operator controls required for generating data and initiating transfers.
- The incremental printer operates at 30 characters per second and provides a hard

copy of console messages. The cabinet containing the printer also contains the power supplies and control logic required to select the CRT, incremental printers, and facilities for the real-time maintenance communication system (RTMCS). This unit also contains the interface between the console and any ISI channel on the IOAU. Up to five additional incremental printers may be connected to the console.

- The fault indicator, located on the incremental printer, provides the operator with a visual indication of a fault condition in a major system component. The actual component and nature of the fault may then be determined from indicators on the maintenance panel.

The RTP System includes one console consisting of one CRT and one printer.

System Partitioning Unit (SPU)

The system partitioning unit (SPU), when included in the UNIVAC 1110 System, permits off-line maintenance of units, enables the operator to logically partition the system into two or three independent systems, and initiates a recovery sequence in the event of failure. The SPU performs six functions, five under operator control and one under software control. With the SPU, the operator can:

- Partition the total system into two or three smaller systems
- Isolate units and take them off-line for maintenance without disrupting the rest of the system
- Function as a system monitor by observing the status of the various major components
- Perform initial load into the primary system
- Allow automatic recovery procedures if an interrupt is not received.

Under software control, the SPU presents status information to the IOAUs. When all optional features are included, the SPU is able to interface with:

- Six command/arithmetic units
- Four IOAUs
- 262K words of primary storage
- Eight MAI units (1048K words of extended storage)
- 48 multiaccess subsystems.

This unit is not included in the RTP System. Since the RTP System has only one IOAU, the SPU cannot be utilized.

Primary Storage

The first level of directly addressable main storage in the UNIVAC 1110 System is primary storage. Primary storage consists of high-speed, nondestructive readout (NDRO) plated-wire storage units with nominal random read and write cycles of 320 and 520 nanoseconds, respectively. The basic 32K-word storage unit consists of four 8K modules, and may be expanded in one 32K increment to a 65K unit. The minimum primary storage for a basic 1x1 configuration (one CAU and one IOAU) consists of 32K words. A total of four 65K units provides a maximum primary storage capacity of 262K words in a system. The basic 32K storage unit accommodates eight access paths, servicing four of them simultaneously; a 65K unit accommodates up to sixteen access paths, servicing eight of them simultaneously. Partial (sixth, quarter, third, and half) as well as full-word operation is provided. The RTP System contains 97K words of primary storage.

Extended Storage

The second level of directly addressable main storage in the UNIVAC 1110 System is the extended storage system. The minimum extended storage configuration consists of 131K 36-bit words. Extended storage capacity may be expanded in 131K increments up to a maximum of 1048K 36-bit words. Each unit has a 1.5 microsecond read/write cycle. Extended storage is connected to the system by MAI units which provide up to ten access paths to each storage unit. The RTP System supports 262K words of extended storage.

Maintenance Controller

The maintenance controller provides for diagnostic checkout by the automatic comparison of maintenance

panel indicators against known good data on tape for the following:

- CAUs
- IOAUs
- Disc controllers (UNIVAC 8440 Disc Subsystem)
- Communication/symbiont processor
- Printed circuit cards.

To complement its diagnostic capability, the maintenance controller allows for the operation of the operator/ maintenance panels by personnel at a remote site. This device is included in the TRP System.

Peripheral Subsystems

The UNIVAC 1110 System offers a full range of peripheral subsystems; this wide range provides the capability to satisfy many requirements. The standard UNIVAC peripheral subsystems include:

- High-Speed Printer Subsystem
  RTP System: 2 High-Speed Printers
- Card Subsystem
  RTP System: 2 Card Readers
- UNIVAC 9000 Series Subsystem
  RTP System: 2 9300 Subsystems, each with 600 LPM Printer
- UNISERVO 12/16 Magnetic Tape Subsystem
  RTP System: 15 Uniservo 16 Tape Drives
- UNISERVO 20 Matnetic Tape Subsystem
  RTP System: None
- FH-432/1782 Drum Subsystem
  RTP System: 6 FH432 Drums, 2 FH1782 Drums
- UNIVAC 8414 Disc Subsystem
  RTP System: None
- UNIVAC 8424/8425 Disc Subsystem
  RTP System: 1 8424 Disc Subsystem (8 Drives)
- UNIVAC 8440 Disc Subsystem
  RTP System: None
- UNIVAC 8460 Disc Subsystem
  RTP System: 4 8460 Discs
- Communications/Symbiont Processor (C/SP)
  RTP System: None
- UNIVAC DCT 500 Data Communications Terminal
  RTP System: 25 DCT-500s are included
- UNIVAC DCT 1000 Data Communications Terminal
  RTP System: None
- UNISCOPE 100 Display Terminal
  RTP System: 10 Uniscope 100s are included
- Communications Terminal Module Controller (CTMC)
  RTP System: 2 CTMCs are included.

Destandardized Subsystems

The following peripheral subsystems may be included in the UNIVAC 1110 System:

- UNISERVO VIIIC Magnetic Tape Subsystem
- FASTRAND II and III Mass Storage Subsystems
- Communication Terminal Synchronous (CTS)
- UNIVAC 1004 Subsystem
- UNISCOPE 300 Visual Communications Terminal.

CONFIGURATIONS

The basic UNIVAC 1110 Processing System (1x1 configuration) consists of two functionally and physically independent units: one CAU and one IOAU. The processor organization is intrinsically that of a multitask processor and is designed for operation in a multiprogramming and multiprocessing environment. The basic processor may be expanded by adding CAUs and/or IOAUs up to a total of four CAUs and four IOAUs (4x4). The basic 1x1 configuration is shown in Figure 2; Table I lists all fully supported configurations.

Table 1
Fully Supported Configurations

UNITS	CONFIGURATION				
	1x1	2x1	2x2	4x2	4x4
CAU	1	2	2	4	4
IOAU	1	1	2	2	4
PRIMARY STORAGE (words)	32K– 262K	65K– 262K	65K– 262K	131K– 262K	131K– 262K
EXTENDED STORAGE (words)	131K–1048K	262K–1048K	262K–1048K	262K–1048K	262K–1048K
MAI	1–8	2–8	2–8	2–8	2–8
SYSTEM CONSOLE	1	1	1–2	2	2–4
SYSTEM PARTITIONING UNIT	0–1	0–1	0–1	1	1

Minimum Peripheral Complement

The following list of peripheral equipment is the minimum available with the UNIVAC 1110 System. This minimum has been established to ensure an adequate complement for customer engineering and software support.

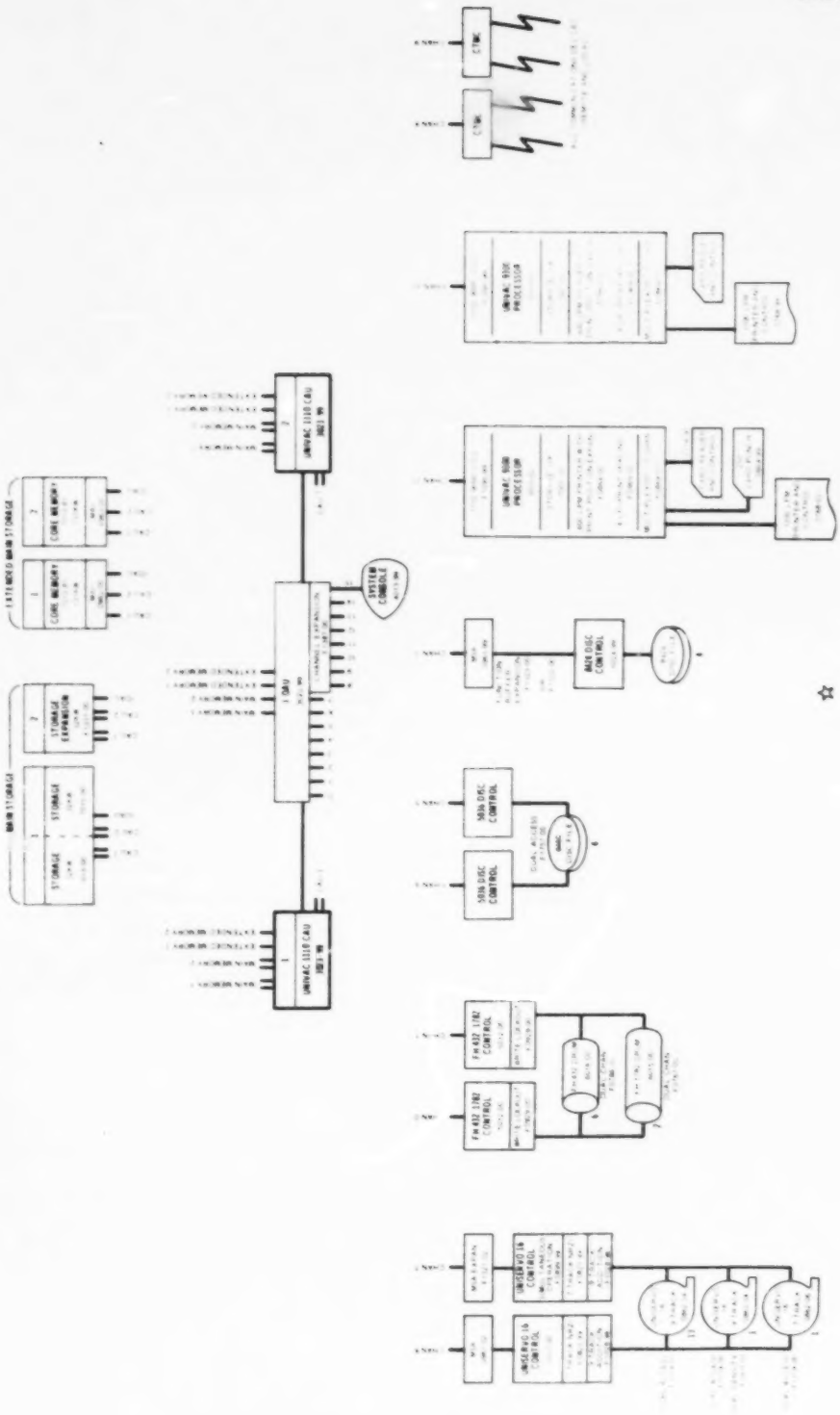
RTP COMMUNICATIONS SUPPORT

The UNIVAC Communications Terminal Module Controller (CTMC) Subsystem enables the UNIVAC 1110 System to receive and transmit data by way of any common carrier at any of the standard rates of transmission up to 50,000 bits per second. It can receive data from or transmit data to low-speed (up to 300 bits per second), medium-speed (up to 1800 bits per second), or high-speed (2000 to 50,000 bits per second) lines in any combination. The RTP system contains two CTMCs giving a total of 64 ports for terminal activity. Figure 3 shows how these 64 ports are configured on the RTP system.

Minimum Complement	Alternate
1. Communications/Symbiont Processor (C/SP) with card reader and high-speed printer	UNIVAC 9300 Subsystem with card reader and integral printer, or Multisubsystem Adapter with high-speed printer subsystem and card subsystem
2. Drum Subsystem*	
FH–432/1782 Drum Subsystem with two FH–432 drums	FH–432/1782 Drum Subsystem with one FH–1782 drum
3. Mass Storage Subsystem	
UNIVAC 8414 Disc Subsystem with two 8414 disc drives	UNIVAC 8424/8425 Disc Subsystem with two 8424/8425 disc drives, or UNIVAC 8440 Disc Subsystem with one 8440 disc drive, or UNIVAC 8460 Disc Subsystem with one disc file unit
4. Magnetic Tape Subsystem	
UNIVAC 12/16 Magnetic Tape Subsystem with four magnetic tape units	UNISERVO VIIIC Magnetic Tape Subsystem with four magnetic tape units

*Not required for disc resident systems (1x1, 2x1, and 2x2)

Figure 1
Central Site Configuration
UNIVAC 1110 System



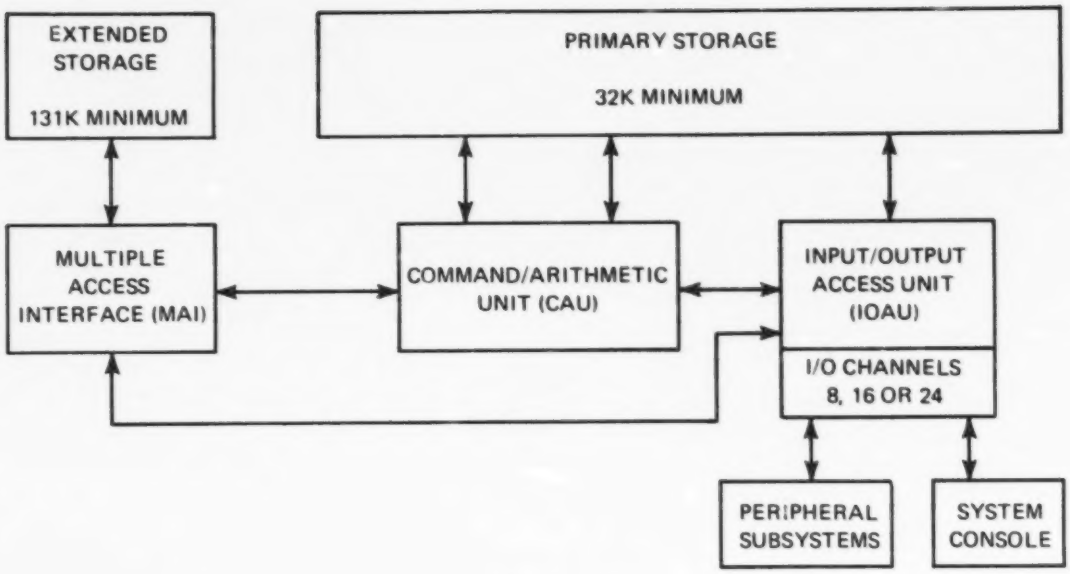


Figure 2
UNIVAC 1110 System
Basic Configuration (1x1)

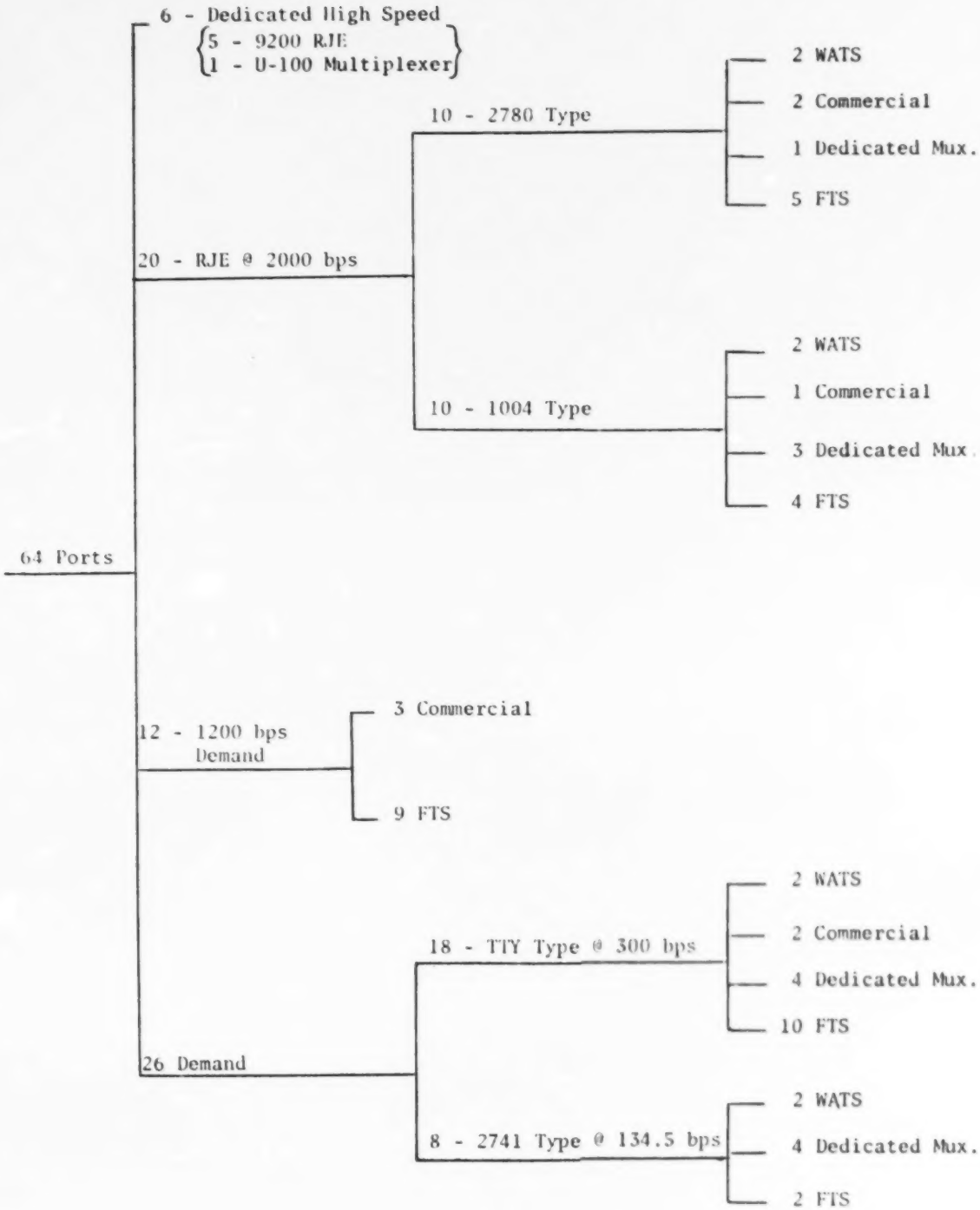


Figure 3
Data Communications Support

RTCC SOFTWARE AND ACCESSIBILITY

By Maureen Johnson

Research Triangle Park offers a full range of data processing software on the UNIVAC 1110 including scientific, statistical and data management packages as well as many compiler languages. This software falls into three general categories:

1. UNIVAC supplied and supported

2. Packages obtained from private vendors or other government agencies

3. Packages converted from the IBM/360 system replaced by the UNIVAC 1110.

1. UNIVAC SOFTWARE

The following is a brief description of available UNIVAC supported software.

1100 Series Assembler

The 1100 Series Assembler contains the following features:

- Mnemonic codes describe hardware function of each instruction
- Multiple location counters provide for program segmentation and control address generation.

ASCII COBOL

Based on American National Standard plus CODASYL and UNIVAC extensions, ASCII COBOL contains the following features:

- STRING and UNSTRING statements providing powerful character manipulation
- Multiple data formats:
  - ASCII
  - FIELDATA
  - EBCDIC–Reading and writing IBM files single and double precision floating point binary
  - Variable length DISPLAY items
- MONITOR and EXHIBIT (MONITOR tremendous aid in program debugging)
- Cross-reference of all data and paragraph names
- Index-sequential file organization on mass storage (some differences from IBM)
- Reentrant object code and libraries, i.e., one copy in memory accessed by multiple users concurrently
- Interprogram and interlanguage communications
- Internal sort capability

FORTRAN V

FORTRAN V, containing all features of American National Standard FORTRAN V plus extensions, has the following characteristics:

- Up to seven subscripts on variables
- Extended subscript expressions NAME (I+1)
- Forward and backward Do Loops
- FLD intrinsic function: used for extraction and insertion of list fields, i.e., bit manipulation
- NAMELIST: may be used instead of a LIST on an INPUT/OUTPUT (I/O) statement and associated FORMAT statement; provides data characteristic information
- DELETE: provides facility to prevent compilation of a section of source code
- Free field input FORMAT ( )
- NTRAN: reads and writes blocks of data
- ERTRAN: means of dynamically executing ECL from FORTRAN program.

ASCII FORTRAN

RTCC is a designated test site for UNIVAC ASCII FORTRAN. This implies that the compiler is available for user testing with the recognition that it is still in a prerelease stage and compiler problems are likely to be encountered. It features:

- ASCII file formats compatible with ASCII COBOL
- List-directed I/O
- Enhanced debugging capabilities including TRACE & DISPLAY.

STAT-PACK AND MATH-PACK

Comprehensive libraries of statistical and mathematical subprograms callable from FORTRAN (comparable to IBM Scientific Subroutine Package).

The STAT-PACK subprgrams are grouped into 13 categories:

- Descriptive statistics
- Elementary population statistics
- Distribution fitting and plotting
- Chi-square tests
- Significance tests
- Confidence intervals
- Analysis of variance
- Regression analysis
- Time series analysis
- Multivariate analysis
- Distribution functions
- Inverse distribution functions
- Miscellaneous.

The MATH-PACK subprograms are grouped into 14 categories:

- Interpolation
- Numerical integration
- Solution of equations
- Differentiation
- Polynomial manipulation
- Matrix manipulation: real matrices
- Matrix manipulation: complex matrices
- Matrix manipulation: eigenvalues and eigenvectors
- Matrix manipulation: miscellaneous
- Ordinary differential equations
- Systems of Equations
- Curve fitting
- Pseudo-random number generators
- Specific functions.

Conversational Time Sharing (CTS)

Conversational Time Sharing (CTS) features the following:

- Extensive file creation and editing capabilities
- FORTRAN prescan capability
- Desk calculator facilities enabling user to evaluate expressions and mathematical functions without any programming.

Functional Mathematical Programming System

The Functional Mathematical Programming System (FMPS) has the following characteristics:

- Procedures commonly used to solve linear programming problems
- Generalized matrix generator
- Procedures for saving the basis, restoring the basis, and the procedures for obtaining error estimates and sensitivity analysis on the solution.

SORT/MERGE

The standalone, parameter-driven sort/merge processor features the following:

- Will process ASCII COBOL fixed length tape records
- Will process FASTRAND-format mass storage card files
- Will process FORTRAN V formatted (80 character) recorded and unformatted files.

PERT

PERT is a generalized applications program for project/program planning and control. It contains both TIME and COST modules.

COBOL Language Conversion Programs

COBOL language conversion programs include:

- IBM PL/1 to UNIVAC ASCII COBOL
- IBM ANSI and COBOL F to UNIVAC ASCII COBOL.

Symbolic Stream Generator

The symbolic stream generator creates a symbolic stream of data and/or control statements with great flexibility and powerful modification capabilities. Directions and models for building stream images are written in SYMSTREAM, an extensive manipulative language.

CPDMPH

CPDMPH is a utility which can be used to print, punch and copy tape or mass storage files.

TAPETRAN

TAPETRAN translates tapes written on other operating systems (primarily EBCDIC oriented) to tapes which are UNIVAC 1110 compatible.

ED

ED is a processor with powerful capabilities in creating and modifying data and program files and elements. It features:

- FIND subcommand for scanning entire file for particular character string
- CHANGE subcommand for modifying one or all entries in a file.

DOC

DOC is a processor for creating, maintaining and retrieving text-type data files.

TPD

TPD provides the facility to do a Directory Listing of a tape or dump it in Alpha or Octal format. Fastran files or elements can also be used as input.

2. PACKAGES OBTAINED FROM PRIVATE VENDORS OR OTHER GOVERNMENT AGENCIES

SPSS

SPSS is a Statistical Package for the Social Sciences and features:

- Flexibility in format of the data
- Routines commonly required by the social scientist including:
  - Descriptive statistics
  - Frequency distributions
  - Simple and partial correlation
  - Multiple regression
  - Factor analysis
  - Guttman Scaling
  - Cross tabulation
- Does not require programming experience.

CALCOMP Graphic/Plotting Routines

CALCOMP Graphic/Plotting Routines are basic plotter subroutines including:

- PLOT
- SYMBOL
- WHERE
- PLOTS
- NUMBER
- SCALE
- AXIS
- FACTOR
- NEWPEN
- LINE
- THREE-D
- GPCP.

A Tektronix interface providing interactive graphics capability has also been developed.

SYMAP

SYMAP is a graphics system designed for making presentations to nontechnical people and has the following characteristics:

- SYMAP uses 10 intervals plus high and low to present the information as a shaded map produced on a printer
- The symbol printed can be changed and the interval can be controlled in such a way that shading density does not increase with increasing values of the variable
- Legends can be included.

SYSTEM 2000

SYSTEM 2000 is a general purpose data base management system and features:

- Creating and modifying data base definitions
- Highly selective and flexible capabilities for retrieving and updating values in these data bases
- Interactive access capabilities
- Report Writer
- Program Language Interface, i.e., user-written COBOL and FORTRAN programs can access the data base.

APL

APL is an interactive computer implementation of a language defined by Kenneth Iverson and named "A Programming Language." The version implemented at RTCC was obtained from the University of Maryland. This implementation is designed to be as nearly like IBM APL/360 as possible.

SNOBOL/SPITBOL

SNOBOL/SPITBOL is a string manipulating language especially well-suited to the processing of non-numeric data. It provides a means for searching through arbitrary character strings in order to find patterns, to rearrange the strings and form new strings.

PL/1 Subset

PL/1 Subset was developed at the University of Maryland and is only available on a testing basis. UNIVAC plans announcement of a PL/1 compiled during the first quarter of 1975.

3. SOFTWARE CONVERTED FROM PACKAGES AVAILABLE ON THE IBM/360

Statistical Analysis System

The Statistical Analysis System (SAS) features:

- Ease and flexibility in defining input data
- Wide range of available statistical routines
- Simple command language requires no programming knowledge
- Data management capabilities.

Time Sharing Library - Interactive Statistical Routines

The Time Sharing Library (TSL) is composed of interactive statistical routines and includes the following characteristics:

- Allows user to enter data and retrieve results through interactive terminals

- Requires no programming experience
- Particularly well-suited for obtaining quick results on relatively small amounts of data.

Keyword In Context

Keyword in Context (KWIC) provides powerful indexing and retrieval capabilities for text storage applications.

FAST

FAST provides the ability to retrieve selected portions of information from a data file and print it in a usable format and/or make it available for further processing. The control cards required are quite simple allowing nondata processing oriented people to use FAST.

ACCESSING THE RTCC UNIVAC 1110

Runs submitted for processing on the UNIVAC 1110 require a valid account number and project code. New users must register via forms provided by Users Services (FTS 919-549-2501) or found in the RTCC Users Reference Manual. The information needed includes user's DIPS Organization Code and Project Element Code, user's mailing address, brief project description and facilities requirements, and a project code made up by the user for each project being registered.

To obtain SITE-IDs for remote and demand terminals, contact Users Services. An information form must be completed describing the terminal, its location and indicating the person to be contacted in matters concerning that terminal. Data Systems will then complete the form with a SITE-ID and phone numbers to be used, and mail a copy to the user.

RTCC SCIENTIFIC SOFTWARE AVAILABLE

RTCC scientific software available includes the following:

- CALCOMP Graphics
- SAS - Statistical Analysis System
- SPSS - Statistical Package for the Social Sciences
- STAT PACK - UNIVAC supplied subprogram callable from FORTRAN
- MATH PACK - UNIVAC supplied subprogram callable from FORTRAN
- TSL - Time Sharing Library of interactive statistical routines
- FMPS - Functional Mathematic Programming System, procedures for solving linear programming problems
- SYMAP - Graphics system
- APL - Interactive computer implementation of a language defined by Kenneth Iverson requiring special character set or dygraphs
- FORTRAN V - Fieldata FORTRAN
- ASCII FORTRAN (RTCC is a designated test-site for ASCII FORTRAN).

DATA MANAGEMENT SOFTWARE AVAILABLE

Data management software available includes the following:

- COBOL (ASCII)
- CTS - Conversational Time Sharing
- PERT - Project/Program Planning Control
- Language Converters
  - IBM COBOL to UNIVAC ASCII COBOL
  - IBM PL/1 to UNIVAC ASCII COBOL
- SSG - Symbolic Stream Generator
- CPDMPH - Prints, punches and copies files
- TAPETRAN - Translates EBCDIC tapes to UNIVAC compatible tapes
- ED - Processor for creating and editing data and programs
- DOC - Processor for creating, maintaining, and retrieving test-type data

- TDP - Dumps tapes and FASTRAND mass storage files and elements
- FLUSH - Flowcharting package
- SYSTEM 2000 - Data Base Management System
- SNOBOL/SPITBOL - String Manipulating Language
- KWIC - Key Word in Contest
- FAST - Generalized data retrieval system.

TECHNICAL AND ENVIRONMENTAL INFORMATION SYSTEM

By Donald L. Worley

INTRODUCTION

The gross incompatibility of the software system (CFSS) that supported APTIC with the new Environmental Protection Agency/Research Triangle Park (EPA/RTP) computer system necessitated a thorough evaluation of APTIC's ADP needs and alternative methods of support.

Results of this evaluation clearly indicated that the most cost effective and recommended course of action was to develop a storage and retrieval system designed to meet APTIC's specific needs. The Data Systems Division (DSD) joined with APTIC in the development and implementation of a system.

The following sections of this paper will describe the basic design of the system developed by DSD for support of APTIC's current requirements.

DESIGN GOALS AND BENEFITS

The Technical and Environmental Information System (TENIS) was designed to provide the following features:

1. The support of the activities which APTIC currently performs on a regular batch processing basis including:

- The processing of input data received from Franklin Institute
- The facility to retrieve selected technical or environmental information through the use of descriptor terms and/or specific document numbers
- The use of a dictionary to control the vocabulary of descriptors either in document input or in the searching process
- The postback dictionary function currently in use by APTIC
- Output may be provided in the form of printed reports or formatted magnetic tapes to be processed by the Government Printing Office
- The maintaining of the current concept and form for document storage and line construction. This will enable the continued use of the programs which use the report tape output of CFSS. These programs will require conversion to Univac compatible code
- A remote terminal search capability.

2. The later expansion of the system to provide capabilities which are desirable but not a necessity at this time. This expansion may include but is not limited to:

- A file maintenance capability for updating and deleting all or selected parts of documents
- An extended search system to provide the capability of the Identification File
- Special reports such as Inverted File Statistics, Output=Terms report, and others which are desirable.

3. A system which is developed and programmed usng a high-order language (COBOL) that is an industry standard and is understandable by the user.

The benefits of this approach to APTIC and DSD are numerous but include:

- The removal of agency dependency on an outside contractor for diagnostic, maintenance and development support
- The reduced developmental costs of a government-produced system
- A system developed specifically for the needs of APTIC
- A system developed by the people who are responsible for using and maintaining it
- The reduced operational cost of a system which meets the specific needs of APTIC
- A modular system which will be expandable as needs dictate.

TENIS SYSTEM DESIGN

An overview of the initial TENIS is shown in Figure 1. The Secondary Search File (previously referred to as the Identification File) is shown by dotted lines as an extension. With the files and systems (groups of programs) reflected on the overview as a reference, a general description of the system follows. A more detailed systems flow chart of programs is included as Attachment A and a sample run of the terminal system is included as Attachment B.

1. *New Document Input.*—Input to TENIS will be structured in the same format as that presently received from Franklin Institute. In addition, the structure and content of documents and lines of text will remain consistent with the present design.

2. *Dictionary.*—The dictionary will be organized as a vocabulary control file, and as an indexer for entering the search file. The postback capability currently provided will be available. One listing of dictionary contents is included as well as the ability to add and delete terms. The ability to delete terms will be provided when the file maintenance extension is implemented. The dictionary will be supported on a direct access device.

3. *Table of Contents.*—A one bit per APTIC Number (1 to 200,000) file to indicate the status of the number. This file shall reside on a direct access device.

4. *Text File.*—The text file will contain data types 01, 035, 036, and 045. One author (06), the journal (09), and the year (10) will be included for sorting output. The text file is designed for efficient use in a batch environment and will be sequential file stored on magnetic tape. Data items are summarized in Figure 2.

5. *Search File.*—The search file shall include all the descriptors necessary for searching. Data types include:

ID No.	Data Item
05	Method of support
10	Year of publication (for internal storage and for searching, it will be prefixed by a P)
13	Language
14	Translated (yes only)
15	Primary category
17	Secondary category
18	Document attributes
19	Indexing descriptors

All entries to the search file must first pass a vocabulary check by the dictionary. The search file will also be directly accessible.

6. *Secondary Search File (not implemented).*—The secondary search file will include the additional descriptors which do not require a vocabulary check. The file will include authors (06,07) and journals (09).

7. *Search Input.* The search input shall use an 80 column card format that is shown in Figure 3. Columns 1 to 4 are fixed and other columns are free form with punctuation.

8. *Document List.* This is a temporary file used in the text retrieval process. It will consist of APTIC identification numbers of all documents that have met the search criteria.

9. *Report Tape.*—The report tape consists of all selected text material (data types 033, 036, and 045) that has been sorted into appropriate reporting sequences. The format will be consistent with the present report tape and may be used as input to many of the current APTIC programs written in COBOL.

10. *Selected Documents.* Documents that have been searched and retrieved according to the pre-entered criteria will be formatted and printed according to present standards.

11. *Edit and File Update System.*—This subsystem consists of six individual COBOL programs and several utility sorts.

12. *Search System.*—These two COBOL programs will provide the Boolean logic necessary to effect the descriptor search of the data files.

13. *Text Retrieval System and Report Program.*—This subsystem consists of two new COBOL programs and several of the publishing programs that will be converted from the present APTIC system.

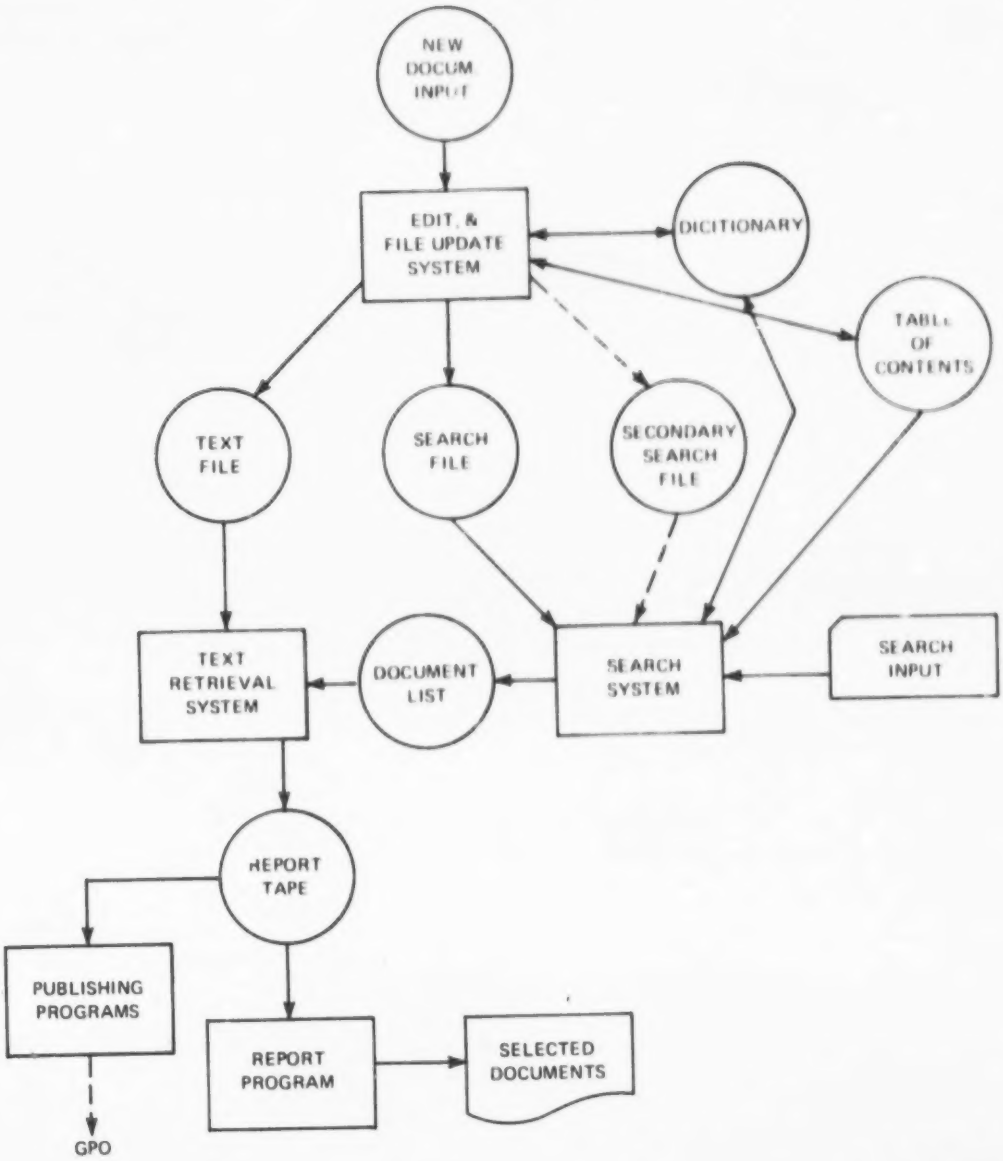


Figure 1
TENIS System Overview

ID	DATA ITEM	MAXIMUM NUMBER	RECEIVING DATA FILE		
			TEXT	SEARCH	SECONDARY SEARCH
01	APTIC Number	1	X	X	X
05	Method of Support	1		X	
06	First Personal Author	1	X		X
07	Other Authors	35			X
09	Title of Publication	1	X		X
10	Year of Publication	1	X	X	
13	Language of Publication	1		X	
14	Translated	1		X	
16	Primary Category	1		X	
17	Secondary Categories				
18	Document Attributes	5		X	
19	Indexing Descriptors				
35	APTIC Number and Authors		X		
36	Bibliographic Citation		X		
45	Abstract		X		

Figure 2
Data Items To Be Accepted
by TENIS

1. General Format

Column 1 to 2 -- A two character search identify number which occurs on every card for a search and which identifies the output.

Column 3 to 4 -- Card Type

010 to 015	Title Card	1 to 6	occurrences
020 to 025	Control Card	0 to 6	occurrences
* 300 to 599	APTIC Number Search	0 to 300	occurrences
* 100 to 299	Boolean Search	0 to 200	occurrences
600	Reserved for Secondary File Search		

*When used as a sequence number, it must be unique and ascending.

Columns 5 to a column containing an @ -- Control Information

Column containing @ to column 75 -- Comments not processed by Search Program

COL.	1 2	3 4 5	6 to @	@ to 76	77-80
USE.	I I	C T	control	comments	Reserved

2. Title Card

From 0 to 71 characters of title information, to be used on printed report. Printed in order of occurrence for up to 3 lines. The title card may not contain comments (if included, they will be printed as part of title).

3. Control Card

Fields are identified by key term and concluded by ';'. Present value is used if not entered.

a. Maximum Number of Documents

Key is M
Form is M=XXXXXX;
Preset is M = 1000;

b. Output Control

Key is P
Form is P = A, B, T, N, D, any combination which is meanful.

D - Text only (eliminate 35 & 36)
A - Segment 35 & 36
B - Segment 35 & 36 & 45
T - Text Record
N - Document Number Only
Preset is P = B

c. Sort Key

Key is S
Form is S=SN, SN, SN;

N=D - Document Number
N=J - Journal Title (Ascending Only)
N=A - Author Title (Ascending Only)
N=Y - Year
S=A - Ascending
S=D - Descending

4 5
Preset is S = DY, AD

Figure 3
Search Control Input Format

d. Date of Entry

D = MMDDYY; everything after the date
D = -MMDDYY; everything up to the date
D = MMDDYY - MMDDYY; between these two dates but not including them
Preset is D = 00 00 00 - 123199

e. Report Number

Key is R
Form is R=NN;
Preset is R=01;

Note: Control cards are continued by a ';'.
If all presets are acceptable, a control card is not required.

4. Document Number Card

Six digit numbers separated by ';'. Space may occur only between ';' and first digit; and leading zeros are not required.
Ranges may be selected by placing a hyphen between two numbers.

Example: 5178;06034; 5178-06034;

5. Search File

Combinations of operators

< - open parenthesis
> - close parenthesis
? - end of search
: - end of Boolean search
& - and
* - or
_ - not (underline)

operands selected from the dictionary
$ - truncation
; - document or year range selection;
Range selections are made using:
EQ - Equal
GT - Greater than
GE - Greater than or Equal
LT - Less than
LE - Less than or Equal

A document range is selected by 'DOC' and year by 'YR'.
The format is:

;DOC GT 5000 LT 10000;
; YR EQ GT;

The second condition is not required.

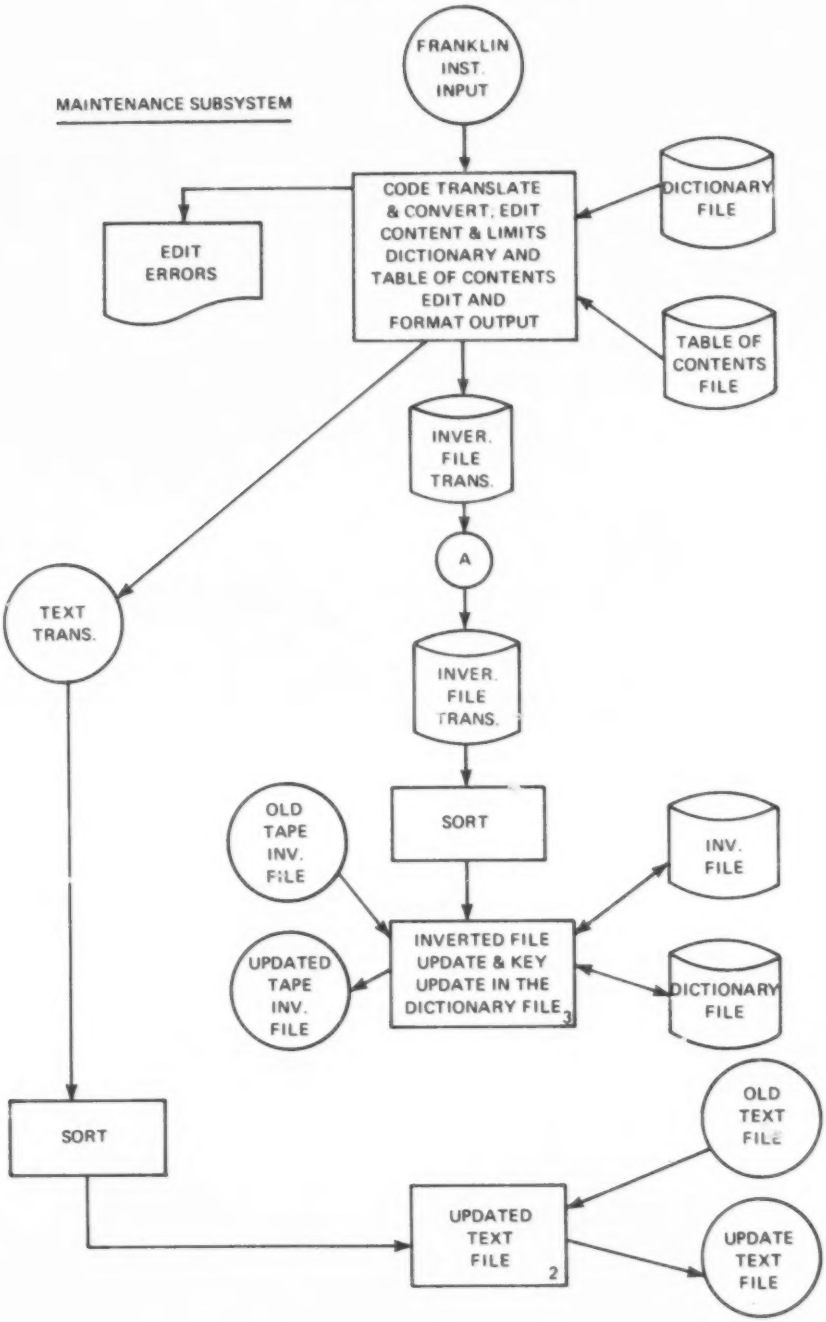
Note: Two range tests may not be combined with themselves.

Acceptable - $R_1 SAS R_2$

Not Acceptable - $(R_1 \; 1 \; R_2) 1A$

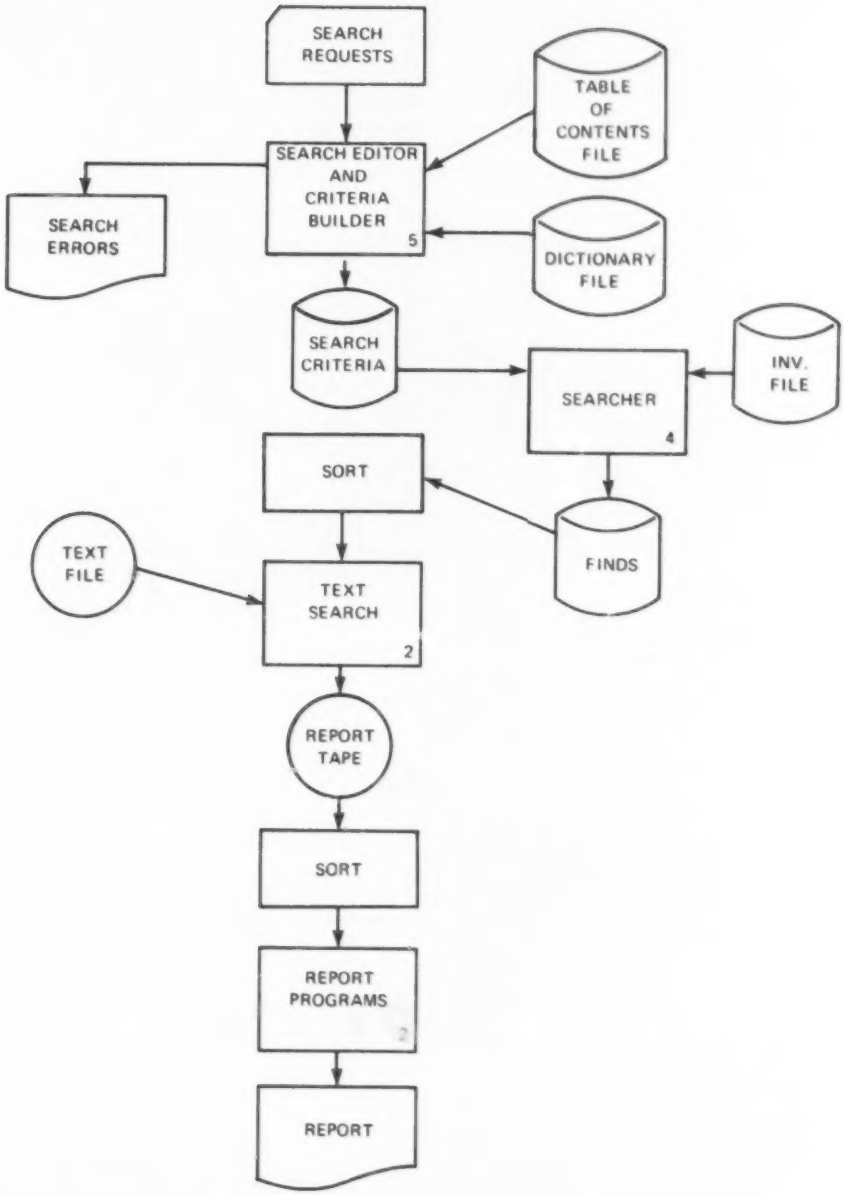
TENIS SYSTEM FLOW CHART

Attachment A(1)

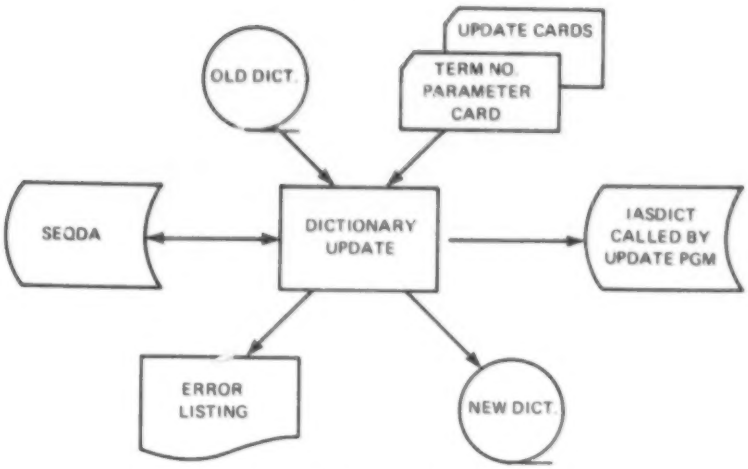


Attachment A(2)

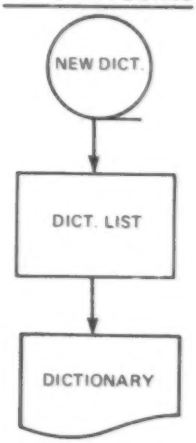
SEARCH AND REPORT SUBSYSTEM



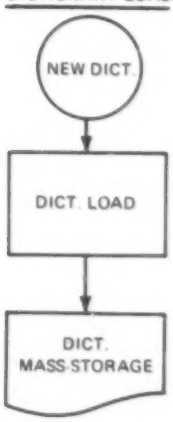
DICTIONARY UPDATE



DICTIONARY LISTING



DICTIONARY LOAD



APTIC Using TENIS

The current implementation of APTIC includes 2500 terms in the dictionary and over 60,000 documents in the master text file. The master file is growing at over 700 documents per month.

Three methods may be used for obtaining results from APTIC:

1. Call Mr. Halpin's group as in the past - and wait for results via mail.

2. Use terminal and review Citation File directly.

3. Use terminal - review Citation File and submit search for results to be returned by mail.

A sample terminal session follows.

```
WELCOME TO TENIS
06 NOV 74     10:54:0
ENTER SEARCH:

>AA010  INITIAL ENTRY
>AA100 ACROLEIN?
      NUMBER OF DOCUMENTS SELECTED = 0  215

INSTRUCT?
>Y
      *ENTER ONE OF THE FOLLOWING
       C TO RECEIVE A CONTINUATION OF DISPLAY OF DOCUMENTS
       G FOR DOCUMENT PRESENTATION, GROUPS OF 5
       B TO RECEIVE SELECTED DOCUMENTS ALSO TYPE
      T TO TERMINATE THE SEARCH
ENTER:
>G
DOCUMENTS NOT DISPLAYED = 0  95

APTIC NO = 0  1 629
      VENGERSKAYA, KH. YA., V. . TARASOV, AND L. I. STARTSEVA
      AIR POLLUTION DURING THE MANUFACTURE OF CHLORINATED OIL. (O
      VREDNYKH VESHCHESTVAKH, YDELYA ISSIK TKSYA V 'OZDUKH PRI
      PROIZVODSTVE KUPROV VIG. MASLA.) SANG. I GIGIENA. (ENGLISH
      TRANSLATION OF: GIGIENA I SANITARIYA.), 33 (1- ):127-130, JAN.-MARCH
      1968.  5 REFS.
         CESTI:  T 68-50049/1

APTIC NO = 0  1 695
      LEONARDOS, GREGORY, DAVID . KENDALL, . D. CY J. REGENAD
      ODOR THRESHOLD ESTIMATIONS FOR 53 ACTIVE CHEMICALS.  PRESENT,
      LITTLE (ARTHUR D.) INC., CAMBRIDGE, MASS., 1968.  4 REFS.
      (PRESENTED AT THE 61ST ANNUAL MEETING OF THE AIR POLLUTION CONTROL
      ASSOCIATION, ST. PAUL, MINN., JUNE 23-28, 1968.  PAPER 68-13.)
```

Reproduced from best available copy.

```
APTIC NO = 0[illegible]1 926
      ALEKSEY VA, M. V., A. S. UZERSKIY, AND V. A. KHRUSTALEVA
      DECREASING THE CONCENTRATION OF TOXIC EXHAUST FROM AUTOMOBILE
      ENGINES. (TO SNIZHENIY KONTSENTRATSIY TOKSICHESKIKH
      OTRABOTANNYKH GAZOV AVTOMOBILI NYKH DVIGATELEY.) TRANSLATED
      FROM RUSSIAN, GI[illegible] (MOSCOW), [illegible] 27:3-7, DEC.
      1962. 1 REF.

APTIC NO = 0[illegible]1[illegible]23
      LITTLE (ARTHUR D.) INC., CAMBRIDGE, MASS.
      RESEARCH ON CHEMICAL ODORS. PART I - [illegible] THRESHOLDS FOR 53
      COMMERCIAL CHEMICALS. MANUFACTURING CHEMISTS ASSOCIATION,
      WASHINGTON, D.C., 26[illegible], OCT. 1968

APTIC NO = 0[illegible]1[illegible]53
      CAMPBELL, KIRBY I., L. OTTS EMIK, GEORGE L. CLARKE, AND
      ROGER L. PLAT
      INHALATION TOXICITY OF THE AIR POLLUTANT PEROXYACETYL NITRATE:
      DEPRESSION OF VOLUNTARY ACTIVITY IN MICE. PRESENT, CALIFORNIA
      UNIV., RIVERSIDE, STATEWIDE AIR POLLUTION RESEARCH CENTER,
      AND PUBLIC HEALTH SERVICE, CINCINNATI, OHIO, NATIONAL AIR
      POLLUTION CONTROL ADMINISTRATION (114)P., (1968). M REFS.


      *ENTER [illegible]
>T
SEARCH TERMINATED
NEW SEARCH? >
ENTER SEARCH:

>AA010.ACROLEIN WITH LIMITING PARAMETER
>AA100 ACROLEIN & ; YR GE 1973 ; ?
      NUMBER OF DOCUMENTS ELECTED = 0[illegible]23

INSTRUCT?
>N
ENTER:
>B
SEARCH SUBMITED TO BATCH COLLECTOR
NEW SEARCH? >QFIN
IN EXEC MODE

  RUNID: TENIS    ACCT: 4502[illegible]  PROJECT: TENIS
M1  BLOCK  SIZE :      0  MERG  POWER :      0
M2  BLOCK  SIZE :      0  MERG  POWER :      0
      SORT:  /=073.0MS
       IC=        3, OC=        3, BIA=SEQ.
SORT:  /=073.0MS, IC=        3, OC=        3, BIA=SEQ.
M1  BLOCK  SIZE :      0  MERG  POWER :      0
M2  BLOCK  SIZE :      0  MERG  POWER :      0
      SORT:  /=921.[illegible]
       IC=        4, OC=        4, BIA=SEQ.
SORT:  /=921.5MS, IC=        4, OC=        4, BIA=SEQ.
  TIME:     [illegible]
            CAU:   0 :  [illegible]2.[illegible]7        I/O:  0 :  [illegible]54.5[illegible]1
            C /FR: 0 :  [illegible]26.193        WAIT: 0 : 7:58.601
  SEC:      PS -   321 [illegible]4              ES -  149 705
  STAR :    10:53:27 NOV 06,1974          FIN:  1 :04[illegible] NOV 06,1974
*TERMINAL [illegible]
>
```


Reproduced from best available copy.

THE CONVERSION OF CHESS AND OTHER SYSTEMS

By Andrea Kelsey, Gene R. Lowrimore,* and Jane Smith

THE CHESS PROGRAM

The CHESS research program is a program of epidemiological studies designed to determine the effects of elevated levels of air pollutants on human health. This program provides the bulk of the information upon which current ambient air quality standards are based. The data processing problem presented by CHESS is the collection and analysis of data on certain disease symptoms, data on demographic characteristics, and data on air pollutant concentrations, from which estimates of the effects of air pollution on health are computed.

Some background on the evolution of the CHESS program will explain the situation that prevailed when the selection of the Univac 1110 system was announced. The CHESS program was conceived about four years ago as a series of replicated studies in several metropolitan areas. When originally planned, these studies were to be conducted independently. Replication was planned from area to area as well as from time to time in the same area. One replication of the study in an area is referred to as a round. Several indicators of human health were measured:

- Chronic Respiratory Disease (CRD)
- Acute Respiratory Disease (ARD)
- Pulmonary Function (PFT)
- Acute Episodes
- Asthma Panel
- Aggravation of Chronic Symptoms (Adult Panels)
- Lower Respiratory Disease (LRD).

By the end of the first year, researchers realized that: (1) ARD panelists should only be selected from the CRD population, (2) data on children in the PFT study should be related to the CRD data for the family from which they came, and (3) any other possible linkages between studies should be made. In this way information could be shared between studies and all available information relevant to the analysis could be obtained.

A short time later an additional extension was made to require us to link together the data of those who participated in more than one round. The extension was made to enable researchers to block out constant person-to-person differences when making comparisons over time. These latter changes in the CHESS program were made in the summer of 1972. Figure 1 shows the relationships between the various CHESS studies for a single round. Partial or total overlapping of clouds implies that data collected on individuals in both studies are linked together. With the exception of CRD, one round of all studies is conducted each year. One round of CRD is done every three years.

The data processing system was redesigned to handle the CHESS concept as it existed in the summer of 1972. Two parallel efforts were thus begun, one to implement the redesigned system, and the other to make past data conform in format and structure to those of the redesigned system.

THE SYSTEMS DESIGN

For the purposes of this paper, the CHESS data processing system is composed of several smaller subsystems, each of which handles one of the health indicators mentioned earlier. A typical subsystem includes processes to:

- Edit background questionnaires
- Establish Master Files
- Print optical mark reader forms
- Read batches of mark reader forms
- Edit (detect errors) data files of periodic responses

*Speaker

- Purge (correct errors) data files of periodic responses

- Combine all periods of responses to update Master Files.

All these operations are standard business-type data processing applications, and they are best done in COBOL. After the Master Files have been updated, they are used in large-scale statistical analyses. Standard packages are used, such as Statistical Package for the Social Sciences (SPSS) and Statistical Analysis System (SAS), along with other user-written packages and routines which are considered less standard. These user-written programs are almost always written in FORTRAN.

Master Files are also used to update what is called a Linkage File. The Linkage File contains all the information collected in all rounds on an individual or a family, as the case may be. This file will also be used for statistical analyses of health status changes over time, but no statistical analyses have actually been done on the Linkage File.

Figure 2 shows an overall system flow indicating the language resources used. Figure 3 is a more detailed description of the subsystem for performing the data processing for the Acute Respiratory Disease indicator. Manual processes are necessary to complete the subsystem but are not shown.

IMPLEMENTATION

Through a work agreement with the General Services Administration (GSA), the Human Studies Laboratory has access to programming and data clerical services furnished by Data Processing Associated (DPA), a GSA contractor. To implement the system, the Laboratory gave DPA a written specification for each process, and DPA delivered a tested program to perform the process along with complete program documentation. The Laboratory maintained control by requiring DPA to obtain approval at several steps in the development process.

The implementation of the redesigned system had been in progress for about seven months when it became necessary to begin the conversion effort. Laboratory project commitments also required the continuation of the development of the IBM version of the system.

CONVERSION PLAN

The following conversion plan was announced concurrently with the procurement announcement of the Univac 1110 computer system:

Step 1: The user would submit source code, systems flow charts and sample run streams for all programs which the user wanted to have converted.

Step 2: The user would submit file descriptions of all files to be converted.

Step 3: The user would submit test data and sample results for the program as run on the IBM 360-50.

Step 4: The contractor would submit complete programs and successful test runs to the user for concurrence.

Step 5: Data Systems Division would convert all user production files. This promise was later withdrawn.

No direct contact between the user and the converting contractor was planned. This procedure was expected to provide the major means for converting user systems from the IBM 360 system to the Univac 1110 system. For reasons discussed below, this plan did not work for the CHESS system. Yet, it seemed to work fairly well for other systems in which the Laboratory has an interest, such as SAS.

CONVERSION, GOOD AND BAD

It has been concluded that contracted conversion for CHESS did not work well for two reasons:

- The CHESS system, like many others, is a changing thing which refuses to remain constant long enough for someone to convert it.

- The users all tend to underestimate how much a system depends upon the environment in which it was developed.

The Laboratory staff submitted about 60,000 lines of source code to be converted under the conversion plan. Between the time of submission of these programs and the time ready for checkout, changed functional requirements, program modifications, and incompatible architectural changes made by the converting contractor,

made it possible to test only about 10,000 lines of source code. The remainder were accepted in the clean-compile state.

After consideration of the above events, the staff became skeptical that "turnkey" conversion is even possible for the majority of systems. Any application system has to be thought of as being embedded in an operating environment. The scope of the system is determined in part by this environment. Capabilities essential to the successful operation of the system are provided by the environment in one operating system but would have to be part of the application system in another. Cases in point are the ability to concatenate files through Job Control Language and the ability to specify blocking factors at the execution time. Both these features are available in the IBM system and absent in the Univac system. Suppose that the files labeled F7 in Figure 3 had been created using several different blocking factors and the files had all been converted using the original blocking factors, then there would be no way that the next process could be run successfully for all these valid inputs. A process also has to be inserted to do the concatenation. In IBM COBOL there is a utility assignment which can be either tape or disk, specifiable at run time. Univac COBOL does not have this capability. The person converting the program is forced to make this decision although it is obviously a systems decision and is greatly influenced by the policy decision made by central site personnel.

Following the decision of the Laboratory staff to perform the conversion of CHESS, the system was redesigned to account for the differences in Univac and IBM environments; these changes were translated into modifications for existing programs and specifications for new programs. No attempt was made to optimize the design of the system with respect to the Univac environment. The staff made all blocking factor and device type decisions and began to convert existing files to Univac media using these factors; they also assigned the conversion of the programs to the Data Processing Associates (DPA) and monitored the effort according to methods employed in any other program development. Since then the conversion of CHESS to the Univac has gone smoothly. Of course, there are problems, such as a bug in ASCII COBOL, which prevents the conversion of the variable blocked files. Comprehensive training was made available to the users by the Data Systems Division and was one of the big factors which enabled the staff to confidently assume the conversion effort.

Some of the problems encountered in the conversion of CHESS are evident even for the conversion of SAS which, except for some delays, has gone well. While the official goal for the conversion of SAS was that all features available in the IBM version would also be available in the Univac version, fairly restrictive requirements were, in fact, placed on input by the converting programmers. The input now must be either cards or a catalogued tape or disk file written one record per block in unformatted FORTRAN. A record may not exceed 3200 characters. The IBM version had a record length restriction of 32000 characters and accepted files with any blocking factors. Also, IBM SAS provided its own sort routine for small sorts and interfaced with the system sort utility for large sorts. The Univac version has no interface with a system sort utility. Of course, many of these restrictions are made inevitable by the Univac operating environment, but had these decisions not been left to the discretion of the converting programmer, there would almost certainly be fewer input restrictions in the Univac version.

INCREASED CAPABILITIES

Increased capabilities resulting from the design changes are as follows:

- Fortran V is more powerful than Fortran IV, particularly in the area of string manipulation. The extra precision provided by the longer word of Fortran V in many cases would obviate the use of double precision.

- Run stream construction is somewhat simpler but may be partially due to the decreased flexibility.

- Management of program files is probably the single most impressive feature of the Univac 1110 system. Source programs, run streams, and executable programs, are all easily established as program files. Powerful software is available to edit these files. Source decks will probably no longer be used except as a means to establish the initial version of a file. Up to five updates of a given file can be kept and easily referenced.

DECREASED CAPABILITIES

Decreased capabilities resulting from the design changes are as follows:

- The Univac System is very weak in utility support such as copy routines for data files and stand-alone sorts. Particularly if one

writes in COBOL and it is not feasible to use internal sort capabilities, programs may be built just to perform the sort steps. There is a file dump program called CPDMPH but it does no formatting and is useful only as a last resort.

- Many decisions which were previously made at run time, must now be made at the time the program is designed. Some of these decisions concern: (1) the number of input files for a given external name, (2) the blocking factor, and (3) the storage device. Fixing blocking factors at program design time will probably be a blessing once the conversion period has passed.

- There is no adequate batch terminal support. Univac batch terminals do not provide any indication that a job was successfully entered. Likewise there is no way to determine through the batch terminal what the status of a job is. The absence of such feedback has hampered the process of learning to use the terminals.

- Files written by different processors are incompatible. In the Univac operating system concept, much of the file management has been delegated to the language processors. Each language processor has its own way of writing files. As a result, users must introduce at least one more process into the converted systems just to reformat the file for statistical programs which follow.

- Input formats of analysis packages are incompatible. The input to SAS is FORTRAN unformatted binary. The input to SPSS is FORTRAN formatted. These two packages are not totally interchangeable in function, and if both are needed, parallel files must be provided.

- Management of data files is inadequate. Putting multiple files on tape, identifying them, and using them seems to be an unreasonably difficult task.

MIXED BLESSING

Using a large fixed mass storage device is a new concept for the Laboratory staff. Instead of having many small files on tape, they may now be stored on mass storage as catalogued files. If the system runs out of disk space, it rolls some of the less frequently referenced files out to tapes, a process which is almost transparent to the user. This concept greatly improves the ability to perform production data processing. But in the process of rolling files in and out, files have been lost. Sometimes the latest good system backup is one week old. The whole concept of file backup has to be reconsidered to keep from permanently losing data.

CHANGES TO CHESS SYSTEM

One of the biggest factors in the development of future systems is the availability of a powerful Data Base Management System, System 2000, on the Univac 1110 System. Establishment of the Linkage File is being considered as well as the Master File shown in Figure 2 as System 2000 Data Bases. Such a move would offer increased flexibility in terms of responding to uncertain requirements for statistical analyses. However, many additional people, mostly analysis programmers, would have to be trained in the use of System 2000 to allow the change to be of any real benefit.

Several other design changes are being considered, such as the on-line input and correction to data.

RECOMMENDATIONS

For similar conversions in the future, the following approaches are recommended:

1. Those systems which are reasonably stable and have a clearly-defined function, such as SAS, can be turned over to a contractor to convert as a system. Even for these systems, the primary communication channel must be between the user and the contractor converting the system. The user must approve all design changes before they are implemented.

2. For other systems, the user must redesign the system to be compatible with the new operating system. A level-of-effort contract can be helpful to convert and modify programs and to convert files. But the important point is that this contract is an extension of the user and not an alternative route for the conversion.

3. Put first priority on making the batch terminals intelligent. This in itself would have a surprisingly positive effect on getting off the old machine and onto the new.

Figure 1
Relationships of the Various Populations for Health Indicators in CHESS

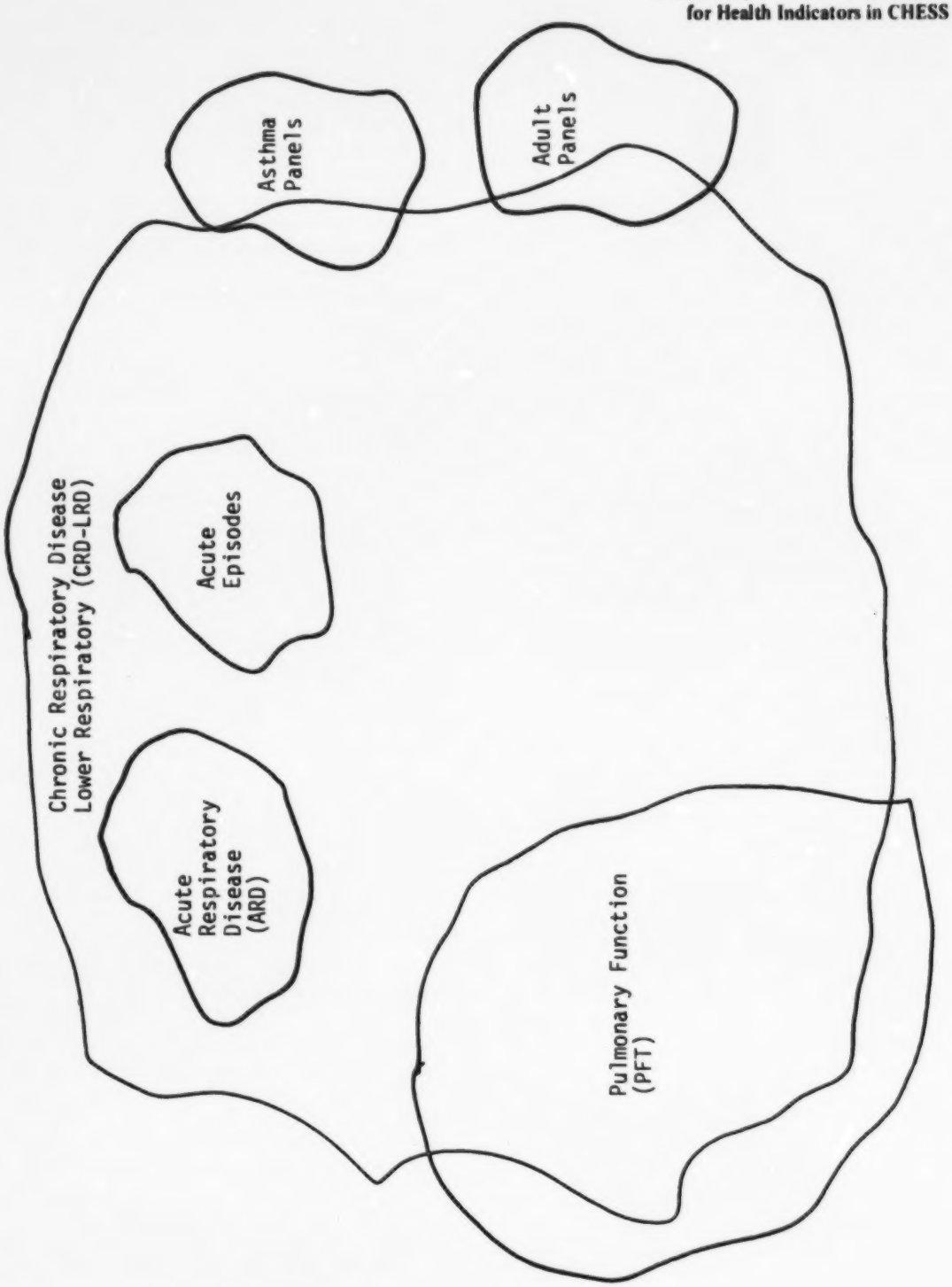
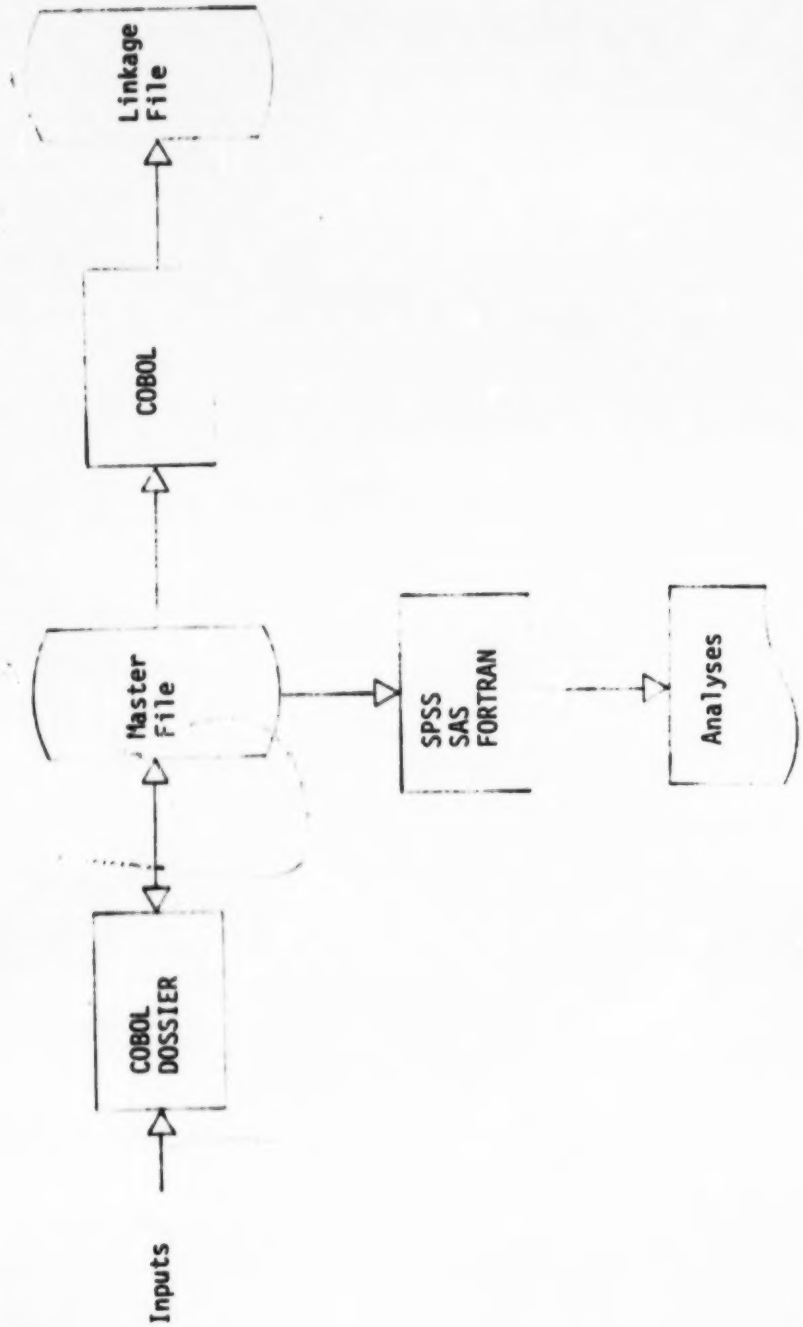
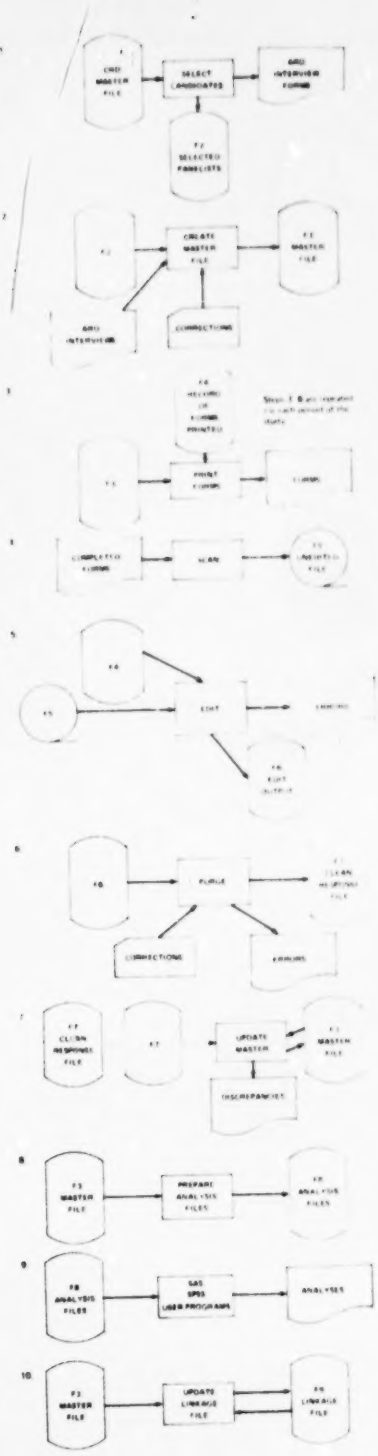


Figure 2
Language Resources for Various Phases of a Typical CHESS Subsystem Showing Major Outputs





1. This first step in the ARD system depends upon the CRD Master File having been completed. The forms printed are readable on a Sentry 70 Optical Mark Reader. Interviewers visit the families selected and attempt to enroll them in the study. Any listing errors in the information about the family is corrected at that time.

2. The Master File has two functions. First it is used to print all interview forms and secondly it will contain all background and response information for a family. Direct access is required because interview forms are randomized and variable length records are used because of the distribution of family sizes.

3. All identifying information is printed on the form and also slugged in machine readable format. Each family is called every two weeks for 32 weeks.

4. The forms are read on a Sentry 70 Optical Mark Reader to 9 track 1600 bpi EBCDIC tape.

5. Logical errors are spotted and recorded in file F6 along with each record somewhat reformatted. There is one record per form.

6. The output F7 is the corrected file and is optional. It is only created after the deck of corrections has been cleaned up and there are no remaining errors which could be corrected.

7. Again the output is optional. If there are any discrepancies one more cycle back through Step 6 is done. All periods are batched together for this process. When the data is as clean as is reasonable, the output file is updated.

8. The records are reformatted to be consistent with standard packages. It is also necessary to compute certain intermediate values.

9. Canned packages are used wherever possible. Some user written analysis programs are usually required.

10. The data for the current round is added to the data for previous rounds. Information for families in the study for more than one year is tied together before storing.

Figure 3
Description of the Acute Respiratory Disease Subsystem

PRESENT STATE OF EDP POLICY AND DEVELOPMENT IN EPA*

By Theodore R. Harris

This article presents some items currently undertaken by the Management Information and Data Systems Division (MIDSD) of the Environmental Protection Agency (EPA).

AGENCY-WIDE EDP PLAN

The House of Representatives in Report No. 93-1120 has given EPA a directive to conduct a detailed Agency-wide study of its ADP requirements and to report its findings to the House and Senate Appropriation committees. No additional ADP equipment is to be purchased until the study is completed. The milestones for the plan are shown below. The completed plan is to be submitted January 2, 1975.

Agency - Wide EDP Plan Development

8/ 7/74	Statement of work for assistance and consultation in development of plan
9/ 1/74	Award of contract
12/11/74	Review draft of EDP Plan from contractor
12/28/74	Delivery of final EDP Plan from contractor
1/ 2/75	Submittal of EDP Plan to the Whitten Committee

WASHINGTON COMPUTER CENTER PROJECT

This project was established to determine the ADP requirements of the Agency. The milestones for the project are shown below.

Washington Computer Center Project†

†† 9/ 2/74	Feasibility Study, first draft
††10/ 1/74	Feasibility Study incorporating Telecommunications, second draft (mail out to members of Task Force)
††10/16/74	Fourth meeting of Task Force (to approve study)
††11/20/74	EPA review and approval of study (Messner and Alm)
1/ 2/75	Whitten approval of Feasibility Study
4/ 1/75	Development of RFP
5/ 1/75	EPA review and approval of RFP
6/ 1/75	GSA review and approval of RFP
7/ 1/75	Issue RFP to public
10/ 1/75	Question and answer period, receipt of proposals
5/ 1/76	Review and evaluation of proposals
8/ 1/76	Cost evaluation and audit of proposals
10/1/76	Selection and award
4/ 1/77	Delivery and acceptance of equipment/service
7/ 1/77	Test and acceptance
11/ 1/77	Conversion of work load (vendor dependent from 6 to 18 months)

INTERIM EDP RESOURCES PROJECT

The purpose of this project is to procure an interim computer resource to fill the gap between Optimum Systems Inc. (OSI) and the Washington Computer Center. The milestones for this project are shown below.

Interim EDP Resources Project

8/ 5/74	Decision to establish project
12/ 1/74	Development of RFP

*Dates have been updated as of December 5, 1974.

†Project has been inactivated. The Interim EDP Resources Project has taken precedence to provide computer resources by July 1975. The Washington Computer Center Project will be activated after conversion to Interim Computer Resource.

††Milestones were not met.

1/ 2/75	EPA review and approval of RFP
2/ 1/75	GSA review and approval of RFP
2/15/75	Issue RFP to public
4/ 1/75	Question and answer period, receipt of proposals
5/ 1/75	Review and evaluation of proposals
6/ 1/75	Cost evaluation and audit of proposals
7/ 1/75	Selection and award
11/ 1/75	Delivery and acceptance of equipment/service
12/ 1/75	Test and acceptance
4/ 1/76	Conversion of work load

STANDARD EDP COMMUNICATING TERMINAL PROJECT

The goal of this project is to standardize types of terminals and procurement of the terminals for the Agency. Preliminary specifications were sent out in June with responses coming back in late July. The milestones are shown below. The current status is October 1 with momentary issue of the RFP.

Standard EDP Communicating Terminal Project

8/15/74	EPA/MIDSD review and approval of RFP
9/24/74	GSA review and approval of RFP
12/18/74	Issue RFP to public
2/ 1/75	Question and answer period, receipt of proposals
3/ 1/75	Review and evaluation of proposals for all Categories I, II, III and IV
4/ 1/75	Selection and award Categories I and II
5/ 1/75	Benchmark period for Categories III and IV
6/15/75	Selection and award Categories III and IV

SUMMARY OF TERMINAL DESCRIPTION

Category	Description
I.A	Low-speed typewriter style Quality impact print Off-line text editing
I.B	Low-speed typewriter style General purpose portable
I.C	Low-speed typewriter style General purpose nonportable
II.A	Display terminals General purpose A/N Display
II.B	Display terminals Graphics
III.A	Medium-speed remote job Entry terminal
III.B	High-speed remote job Entry terminal
IV.A	Remote job entry with concurrent processing capabilities Data entry oriented
IV.B	Remote job entry with concurrent processing capabilities Scientific oriented

EDP DATA COMMUNICATIONS NETWORK

A feasibility study contract was awarded to determine EPA network feasibility. The study has been completed. A common network with a design linking all users to the two major computer resources of the Agency has been recommended.

Figure 1 shows the EPA network WATS telephone service. Figure 2 shows the multiplexor placement. Figure 3 shows a typical multiplexor configuration. Figure 4 shows a functional representation of the network. The milestones of the project are shown below.

EDP Data Communications Network

6/10/74	Date of contract award for "Feasibility of an EPA Data Communications Network"
9/18/74	Delivery of Data Communications Feasibility Study Report
12/ 1/74	Initiation of Phase II, produce RFP specs
12/15/74	Delivery of RFP from contractor
1/ 2/75	EPA review and approval of RFP
1/ 2/75	Whitten approval of feasibility study
2/ 1/75	GSA review and approval of RFP
2/15/75	Issue RFP to public
4/ 1/75	Question and answer period, receipt of proposals
5/ 1/75	Evaluate proposals
6/ 1/75	Cost evaluation and audit of proposals
7/ 1/75	Selection and award
11/ 1/75	Delivery and acceptance of equipment/service
12/ 1/75	Test and acceptance
4/ 1/76	Conversion of work load

Figure 1
EPA Network WATS Telephone Service

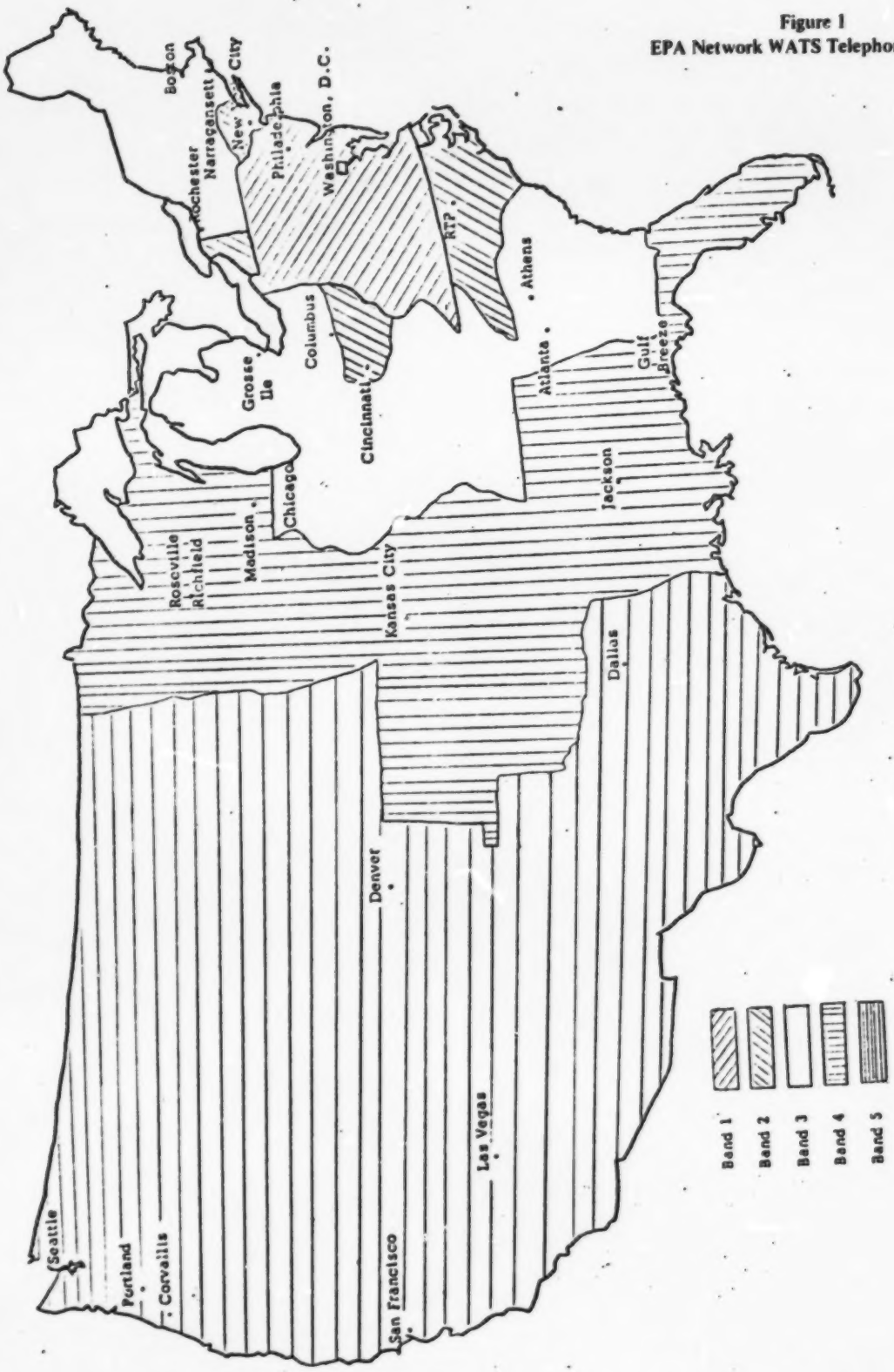


Figure 2
FY 77 EPA Network Multiplexor Placement

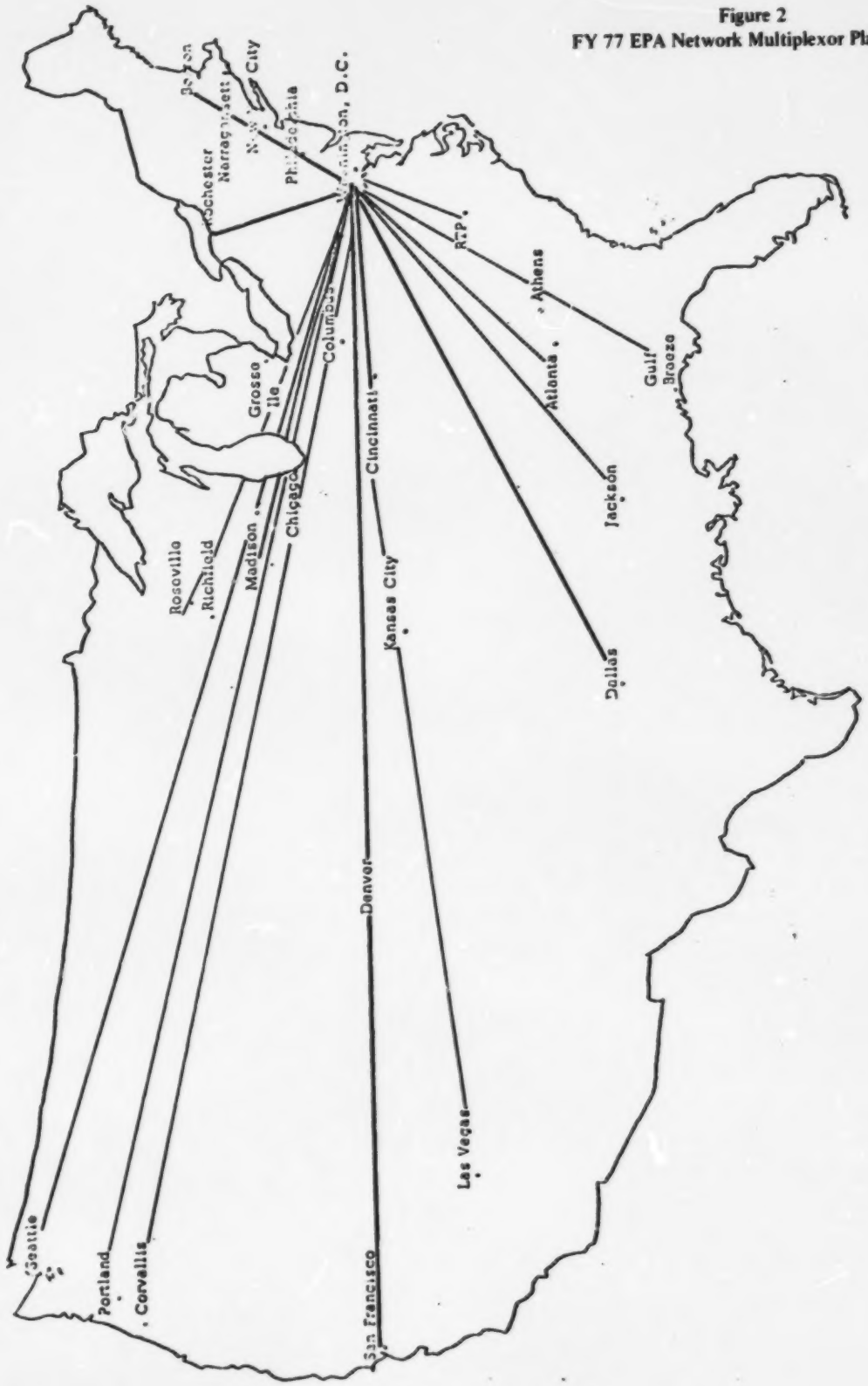
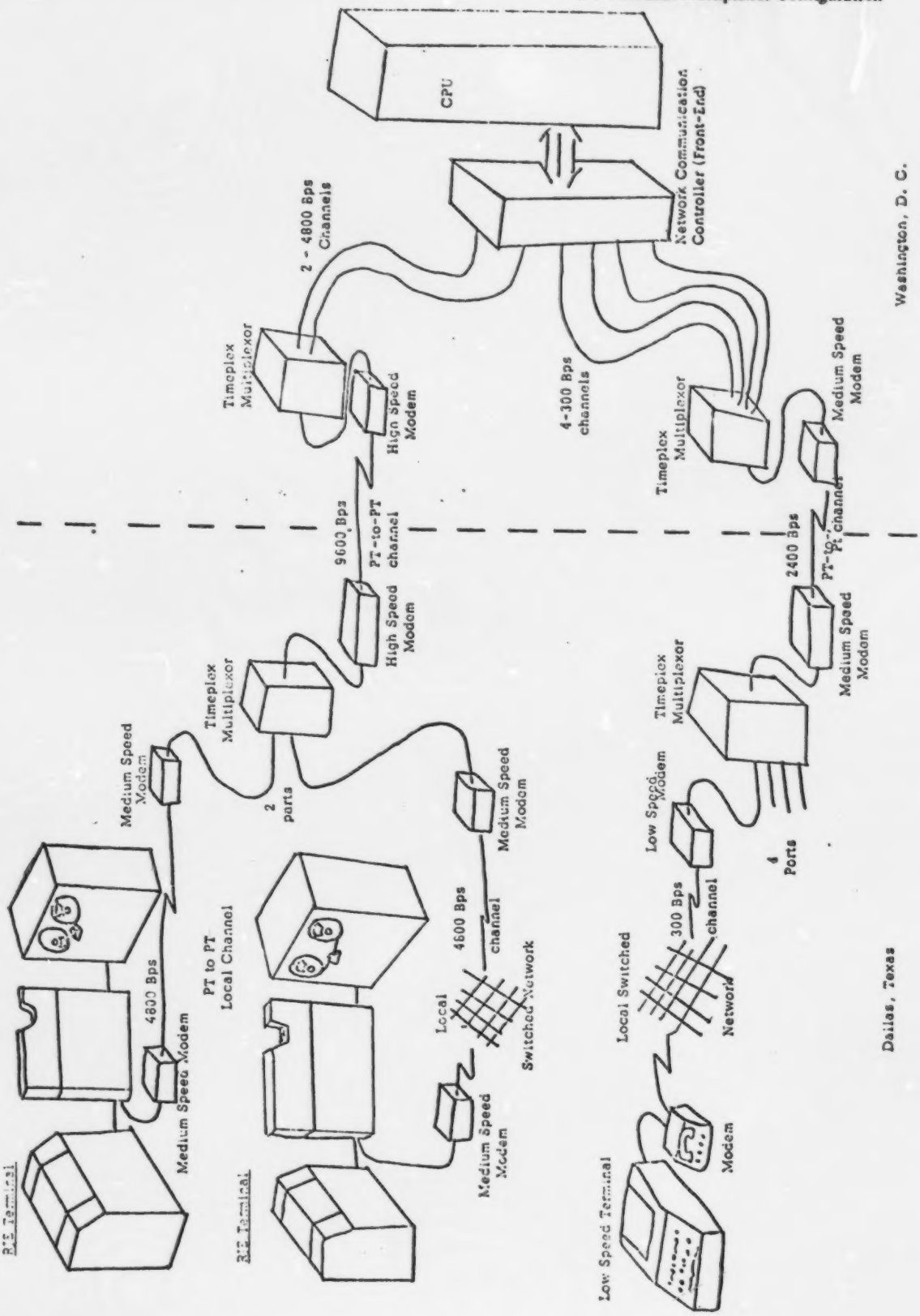


Figure 3
FY 75 Dallas Multiplexor Configuration



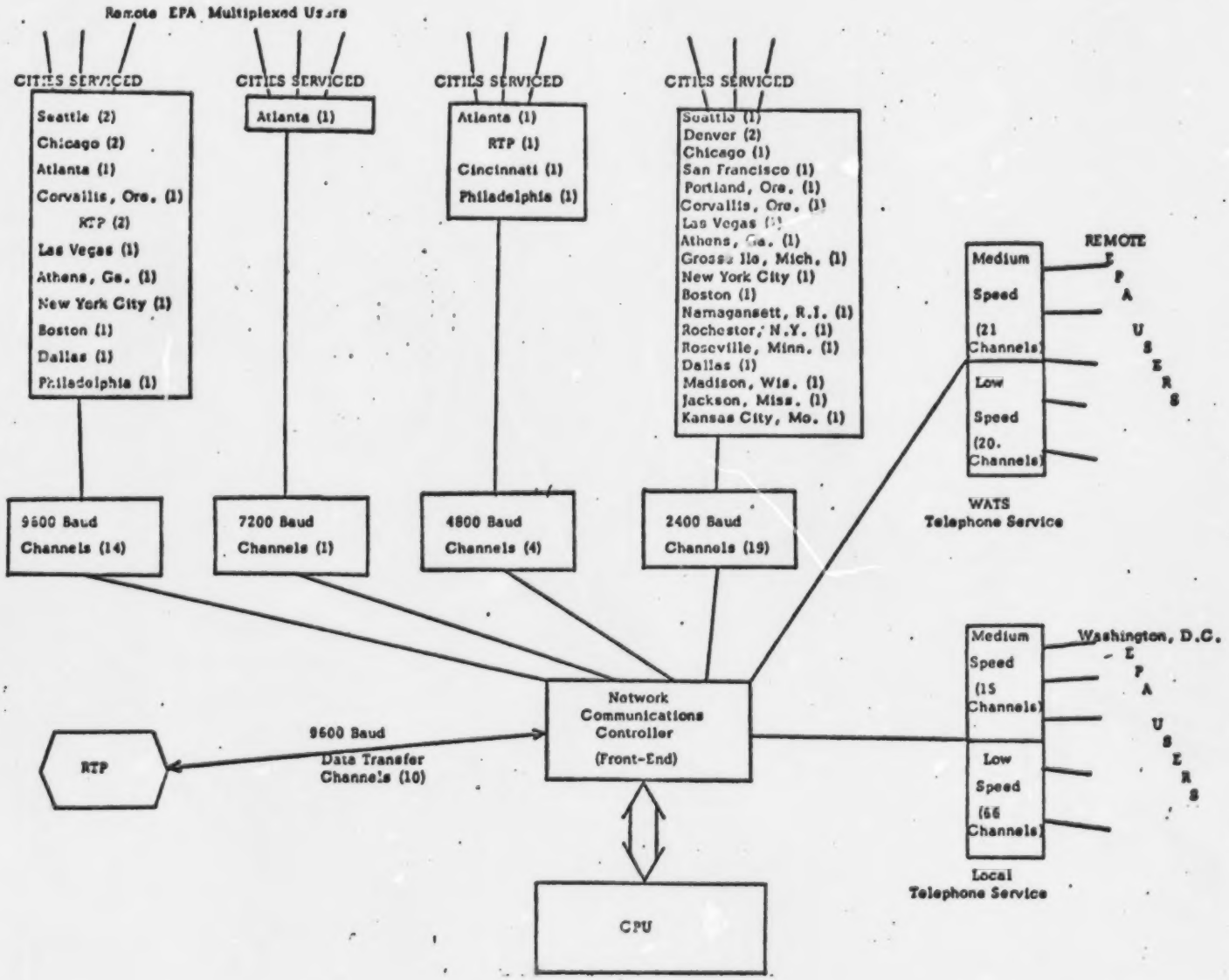


Figure 4
FY 75 Network Configuration

STATUS OF THE WASHINGTON COMPUTER CENTER PROCUREMENT

By Denise Swink

The Washington Computer Center (WCC) Task Force was established by the Office of Planning and Management to evaluate the need for and the resources required to implement and maintain an Agency-wide consolidated ADP facility. The membership of the task force is composed of a representative from each of the Assistant Administrators' and Regional Administrators' offices.

The first meeting of the WCC Task Force was held February 4, 1974, to review the General Electric (GE) Study for content and acceptability. The GE Study defined the Agency's data processing requirements and recommended alternative courses of action to support the requirements. Members were to produce a response to the GE Study by the next meeting in terms of selecting an alternative, planning necessary personnel support for the alternative, and planning for additional ADP funding to support the alternative.

The second meeting of the WCC Task Force was held March 15, 1974. At that time, the consensus of opinion was that the GE Study should be rejected and that a new feasibility study should be developed by the Task Force. Reasons for eliminating the GE Study from consideration and evaluation were based on the facts that the study was out-of-date (two years old at the time) and that the workload information presented in the study was inaccurate. Consequently, a Validation Committee was established to perform an Agency-wide workload survey.

The third meeting of the WCC Task Force was held April 29, 1974. The Validation Committee presented the findings of the workload survey noting that at the time they had only a 50 percent response. A revalidation cycle was suggested by the committee to complete the survey and correct any erroneous data. Michael Springer, Chairman of the WCC Task Force, then established five working groups with functions as follows:

1. General—coordination of other groups' activities and higher-management interaction.

2. Workload, Expansion, and Feasibility Study—production of a formal document (WCC Feasibility Study) which defines the Agency's data processing workload requirements now and in the future, and recommends a course of action to support the requirements.

3. Hardware, Software, and Telecommunications—production of RFP specifications to implement action recommended in the WCC Feasibility Study in the areas of hardware, software, and telecommunications.

4. Facilities Management and Security—production of RFP specifications to implement action recommended in the WCC Feasibility Study in the areas of facilities management and security.

5. Conversion and Benchmark—production of RFP specifications to convert and benchmark the EPA workload on the action recommended in the WCC Feasibility Study.

The fourth meeting of the WCC Task Force hinges on the completion of the WCC Feasibility Study since three of the working groups cannot produce documents until the results of the feasibility study are available.

The Workload, Expansion and Feasibility Study Group began meeting May 16, 1974. After completing the Validation Survey, the group designed, developed, and wrote the Feasibility Study. The group met through July at three-week intervals, five days in August, and four days in September. The final draft of the Feasibility Study is now in the review process by the group. The study includes the following:

- Definition of present services including workload data and reasons for using a particular service
- Definition of problems with present service arrangements
- Definition of requirements
- Definition of feasible alternatives
- Analysis of alternatives (technical, managerial, and cost)
- Recommended action.

It should be noted that the Washington Computer Center was considered a misnomer by the group since there is no requirement that the facility for the implementation of possible alternatives be physically located in the Washington, D.C. area.

The WCC Feasibility Study along with the Telecommunications Feasibility Study (directed by Management Information and Data Systems Division), and the EDP Plan required by Congress will be used by higher management to procure, beginning in fiscal year 1978, the ADP services required to support EPA programs for a life cycle of eight to ten years.

APPENDIX A

LIST OF ATTENDEES

Allen, Ralph G.
Senior Programmer
D. P. Associates

Grosse Ile Laboratory
National Environmental Research Center
9311 Groh Road
Grosse Ile, Michigan 48138

Barrow, David R.
Systems Programmer
Surveillance and Analysis Division

Southeastern Environmental Research Laboratory
Environmental Protection Agency
College Station Road
Athens, Georgia 30601

Bliss, James D.
Environmentalist
Monitoring Operations Laboratory

National Environmental Research Center
Environmental Protection Agency
P.O. Box 15027
Las Vegas, Nevada 89114

Borthwick, Patrick W.
Biologist/ADP Coordinator
Gulf Breeze Environmental Research Laboratory

Gulf Breeze Environmental Research Laboratory
National Environmental Research Center
Environmental Protection Agency
Sabine Island
Gulf Breeze, Florida 32561

Broadway, Jon A.
Supervisor, Computer Services Section
Oil and Special Materials Division, Office of Radiation Programs

Environmental Protection Agency
P.O. Box 3009
Montgomery, Alabama 36109

Brooks, Dorothy
Computer Programmer
D. P. Associates

Grosse Ile Laboratory
National Environmental Research Center
9311 Groh Road
Grosse Ile, Michigan 48138

Bryan, Sam D.
Mathematician
Clinical Studies Branch

Human Studies Laboratory
National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Budde, Bill
Chief, Advanced Instrumentation Section
Advanced Instrumentation Section

Methods Development and Quality Assurance Research Laboratory
National Environmental Research Center
Environmental Protection Agency
Cincinnati, Ohio 45268

Burton, Judy K.
Computer Programmer
Laboratory Services Branch (PNERL)

National Environmental Research Center
Environmental Protection Agency
200 S.W. 35th Street
Corvallis, Oregon 97330

Byram, Kenneth V.
Computer Specialist
Laboratory Services Branch (PNERL)

National Environmental Research Center
Environmental Protection Agency
200 S.W. 35th Street
Corvallis, Oregon 97330

Bystroff, Roman I.
Chemist
Lawrence Livermore Laboratory

Lawrence Livermore Laboratory
P.O. Box 808, L-404, Chemistry Department
Livermore, California 94550

Cline, David M.
Computer Systems Analyst
Southeast Environmental Research Laboratory

Southeast Environmental Research Laboratory
National Environmental Research Center
Environmental Protection Agency
College Station Road
Athens, Georgia 30601

Conger, Charles S.
Chief, Information Access and User Assistance Branch
Monitoring and Data Support Division

Office of Water and Hazardous Materials (AW-453)
Environmental Protection Agency
Washington, D.C. 20460

Dell, Robert
Computer Engineer
Central Regional Laboratory

Region V-CRL
Environmental Protection Agency
1819 Pershing Road
Chicago, Illinois 60609

Dipert, Merlin H.
Chief, Data Systems Branch, Management Division
Data Systems Branch, Management Division

Region V
Environmental Protection Agency
1 North Wacker Drive
Chicago, Illinois 60606

Fairless, William
Deputy Director and Chief, Chemistry Branch
Central Regional Laboratory

Region V-CRL
Environmental Protection Agency
1819 W. Pershing Road
Chicago, Illinois 60609

Feldmann, Richard
Computer Specialist
Computer Center Branch

National Institutes of Health
Division of Computer Research and Technology
Bethesda, Maryland 20014

Florence, Cecil E.
Chief, Data Processing Branch
Data Processing Branch, Management Division

Region VII
Environmental Protection Agency
1735 Baltimore
Kansas City, Missouri 64108

Friedland, Michael J.
Systems Analyst/Programmer
Data Services Branch, (TSL)

National Environmental Research Center
Environmental Protection Agency
P.O. Box 15027
Las Vegas, Nevada 89114

Gangler, James
Programmer/Analyst
Vitro Laboratories

Vitro Laboratories
14000 Georgia Avenue
Silver Spring, Maryland 20910

Goldberg, Neal
Computer Programmer
Ecological Research Branch

National Marine Water Quality Laboratory
National Environmental Research Center
Environmental Protection Agency
P.O. Box 277
West Kingston, Rhode Island 02982

Greaves, John O. B.
Assistant Professor
Southeastern Massachusetts University

Department of Electrical Engineering
Southeastern Massachusetts University
North Dartmouth, Massachusetts 02747

Harris, Theodore R.
Computer Specialist
Management Information & Data Systems Division

Office of Planning and Management (PM-218)
Environmental Protection Agency
Washington, D.C. 20460

Hart, John J.
Chief, Systems Analysis and Programming Branch
Computer Services and System (OA)

National Environmental Research Center
Environmental Protection Agency
5555 Ridge Avenue
Cincinnati, Ohio 45268

Heller, Stephen R.
Computer Specialist
Management Information and Data Systems Division

Office of Planning and Management (PM-218)
Environmental Protection Agency
Washington, D.C. 20460

Hertz, Marvin
Chief, Systems Engineering Section
Bio-Environmental Measurement Branch

Human Studies Laboratory
National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Holsomback, Will F.
Computer Specialist
Surveillance and Analysis Division

Southeastern Research Laboratory
Environmental Protection Agency
College Station Road
Athens, Georgia 30601

Johnson, Maureen M.
Computer Specialist
Data Systems Division (OA)

National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Johnson, Richard
Technical Information Specialist
Air Pollution Technology Information Center

Air Pollution Technology Information Center
Room 255
Chemstrand Building
Research Triangle Park, North Carolina 27711

Jurgens, Robert B.
Physicist
Regional Air Pollution Studies Branch

Meteorology Laboratory
National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Kinnison, Robert R.
Mathematical Statistician
Monitoring Systems Analysis Staff (MSRDL)

National Environmental Research Center
Environmental Protection Agency
P.O. Box 15027
Las Vegas, Nevada 89114

Knight, John E.
Technical Information Specialist
Special Studies Staff

National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Kyle, Kirby D.
Physicist
Bio-Environmental Measurement Branch

Human Studies Laboratory
National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Lackey, Curtis S.
Chief, Data Systems Branch
Data Systems Branch, Management Division

Region IV
Environmental Protection Agency
1421 Peachtree Street
Atlanta, Georgia 30309

Laurie, Vernon J.
Physical Science Administrator
Office of Monitoring Systems

Office of Research and Development (RD-687)
Environmental Protection Agency
Washington, D.C. 20460

Lowrimore, Gene R.
Chief, Data Processing Section
Biometry Branch

Human Studies Laboratory
National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Madsen, Mark J.
Systems Analyst
Data Services Branch (TSL)

National Environmental Research Center
Environmental Protection Agency
P.O. Box 15027
Las Vegas, Nevada 89114

Male, Larry M.
Supervisory Operations Research Analyst
National Ecological Research Laboratory

National Environmental Research Center
Environmental Protection Agency
200 S.W. 35th Street
Corvallis, Oregon 97330

McCarthy, William N., Jr.
Chemical Engineer
Office of Program Management

Office of Research and Development (RD-674)
Environmental Protection Agency
Washington, D.C. 20460

McGuire, John M.
Chief, Chromatography and Mass Spectrometry Section
Analytical Chemistry Staff

Southeast Environmental Research Laboratory
National Environmental Research Center
Environmental Protection Agency
College Station Road
Athens, Georgia 30601

Myers, Melvin L.
Center Staff Officer
Program Coordination Staff

National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Nime, Edward J.
Director, Computer Services and Systems
Computer Services and Systems (OA)

National Environmental Research Center
Environmental Protection Agency
5555 Ridge Avenue
Cincinnati, Ohio 45268

Ott, Wayne R.
Systems Analyst
Office of Monitoring Systems

Office of Research and Development (RD-687)
Environmental Protection Agency
Washington, D.C. 20460

Poole, Elijah L.
Computer Specialist
Management Information and Data Systems Divison

Office of Planning and Management (PM-218)
Environmental Protection Agency
Washington, D.C. 20460

Richardson, William L.
Environmentalist
Large Lakes Branch

Grosse Ile Laboratory
National Environmental Research Center
Environmental Protection Agency

Rockwell, David C.
Acting Chief, Data Management Section
Surveillance and Analysis Division

Region V
Environmental Protection Agency
1 North Wacker Drive
Chicago, Illinois 60606

Rogers, Tommie L.
Chief Hardware and Communications Management Section
Data Systems Division (OA)

National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

Schuk, Walter W.
Electronics Technician
Technology Development Support Branch

EPA-DC Pilot Plant
5000 Overlook Avenue S.W.
Washington, D.C. 20032

Scotton, John W.
Technical Information Specialist
Office of Monitoring Systems

Office of Research and Development (RD-689)
Environmental Protection Agency
Washington, D.C. 20460

Shew, D. Craig
Research Chemist
Subsurface Environmental Branch

Robert S. Kerr Environmental
Research Laboratory
National Environmental Research Center
Environmental Protection Agency
P.O. Box 1198
Ada, Oklahoma 74820

Snelling, Robert N.
Chief, Data Services Branch
Data Services Branch (TSL)

National Environmental Research Center
Environmental Protection Agency
P.O. Box 15027
Las Vegas, Nevada 89114

Swink, Denise
ADP Coordinator
Office of Program Management

Office of Research and Development (RD-674)
Environmental Protection Agency
Washington, D.C. 20460

Teuschler, Jack
Electrical Engineer
Instrumentation Development Branch

Methods Development and Quality
Assurance Research Laboratory
National Environmental Research Center
Environmental Protection Agency
Cincinnati, Ohio 45268

Tiffany, William C.
Computer Programmer
Eutrophication Survey Branch (PNERL)

National Environmental Research Center
Environmental Protection Agency
200 S.W. 35th Street
Corvallis, Oregon 97330

Tittle, Catherine
Manager, Customer Services
Bowne Timesharing, Inc.

Bowne Timesharing, Inc.
1025 Connecticut Avenue N.W.
Washington, D.C. 20036

Uchrin, Christopher
Sanitary Engineer
Surveillance and Analysis Division

Region II
Environmental Protection Agency
21 Stonehenge Drive
Lincroft, New Jersey 07738

Webb, Ronald H.
Senior Account Representative
Bowne Timesharing, Inc.

Bowne Timesharing, Inc.
1025 Connecticut Avenue, N.W.
Washington, D.C. 20036

Williams, Edward R.
Program Manager: Comprehensive Analyses
Forecasting and Analysis Branch

Washington Environmental Research Center (RD-691)
Environmental Protection Agency
Washington, D.C. 20460

Williams, Robert T.
Chief, Waste Identification and Analysis
Waste Identification and Analysis Section

Advanced Waste Treatment Research Laboratory
National Environmental Research Center
Environmental Protection Agency
Cincinnati, Ohio 45268

Worley, Donald L.
Chief Analysis and System Design Section
Data Systems Division (OA)

National Environmental Research Center
Environmental Protection Agency
Research Triangle Park, North Carolina 27711

APPENDIX B

AREAS OF EXPERTISE

Air Monitoring Systems

Hertz, Marvin
Jurgens, Robert B.
Kyle, Kirby D.

Analysis, Design, and Programming

Allen, Ralph G.
Barrow, David R.
Brooks, Dorothy
Bryan, Sam D.
Burton, Judy K.
Byram, Kenneth V.
Cline, David M.
Dipert, Merlin H.
Friedland, Michael J.
Gangler, James
Goldberg, Neal
Knight, John E.
Lackey, Curtis S.
Lowrimore, Gene R.
Madsen, Mark J.
Male, Larry M.
McCarthy, William N., Jr.
Ott, Wayne R.
Tiffany, William C.

Analytical Methods Development

Budde, Bill

Automated Chromatography

McGuire, John M.

Basin Planning

Uchrin, Christopher

CHAMP/CHESS

Hertz, Marvin
Kyle, Kirby D.
Lowrimore, Gene R.

Chemical Structure Searching

Feldmann, Richard
Heller, Stephen R.

CLEVER/CLEANS

Lowrimore, Gene R.

Coordination of EPA's Use of Word/One

Tittle, Catherine
Webb, Ronald H.

Data Base Management

Byram, Kenneth V.
Bystroff, Roman I.
Dell, Robert
Florence, Cecil E.
Gangler, James
Jurgens, Robert B.
Knight, John E.
Lowrimore, Gene R.
Madsen, Mark J.
Rockwell, David C.
Scotton, John W.
Snelling, Robert N.
Worley, Donald L.

Data Storage and Retrieval

Bliss, James D.
Friedland, Michael J.
Holsomback, Will F.
Knight, John E.
Richardson, William L.
Worley, Donald L.

Development of Library Reference Spectra

Heller, Stephen R.
McGuire, John M.

Direction and Coordination of ADP Needs

Byram, Kenneth V.
Harris, Theodore R.
Johnson, Maureen M.
Knight, John E.
Nime, Edward J.
Snelling, Robert N.
Swink, Denise

Effects of Pollutants on Organism Behavior

Borthwick, Patrick W.
Greaves, John O.B.

Engineering

Myers, Melvin L.

ENVIR/EDMPAS

Scotton, John W.

Equipment Utilization

McCarthy, William N., Jr.

Facilities Management

McCarthy, William N., Jr.

Forecasts and Costs of Pollution Control

Williams, Edward R.

Geophysics

Rockwell, David C.

Graphics

Cline, David M.
Feldmann, Richard
Goldberg, Neal
Jurgens, Robert B.
Lackey, Curtis S.
Male, Larry M.

Laboratory Automation

Budde, Bill
Byram, Kenneth V.
Bystroff, Roman I.
Cline, David M.
Deli, Robert
Fairless, William
Goldberg, Neal
Heller, Stephen R.
Laurie, Vernon J.
Nime, Edward J.
Ott, Wayne R.
Teuschler, Jack

Modeling

Dipert, Merlin H.
Hart, John J.
Ott, Wayne R.
Richardson, William L.
Williams, Edward R.

Ocean Monitoring

Uchrin, Christopher

Pollution Monitoring Systems

Broadway, Jon A.
Hertz, Marvin
Jurgens, Robert B.
Kinnison, Robert R.
Kyle, Kirby D.
Scotton, John W.
Uchrin, Christopher
Williams, Robert J.

SAS

Bliss, James D.
Johnson, Richard
Lackey, Curtis S.
Lowrimore, Gene R.

SEAS

Williams, Edward R.

Spectrometry

Budde, Bill
Heller, Stephen R.
McGuire, John M.
Shew, D. Craig

Statistical Analysis

Borthwick, Patrick W.
Byram, Kenneth V.
Dell, Robert
Kinnison, Robert R.
Lowrimore, Gene R.
Male, Larry M.
Ott, Wayne R.
Poole, Elijah L.

STORET

Bliss, James D.
Conger, Charles S.
Friedland, Michael J.
Holsomback, Will F.

Telecommunications Development

Harris, Theodore R.
McGuire, John M.
Rogers, Tommie L.

UNIVAC 1110 System

Johnson, Maureen
Rogers, Tommie L.
Worley, Donald L.

Waste Treatment Research and Analysis

Schuk, Walter W.
Williams, Robert T.

Water Quality Analysis

Dipert, Merlin H.
Hart, John J.
Richardson, William L.

APPENDIX C

AREAS OF INTEREST

ADP Utilization/Management

Byram, Kenneth V.
Dipert, Merlin H.
Florence, Cecil E.
Johnson, Maureen M.
Knight, John E.
Lackey, Curtis S.
Myers, Melvin L.
Nime, Edward J.
Poole, Elijah L.
Snelling, Robert N.
Swink, Denise

Air Modeling

Broadway, Jon A.
Jurgens, Robert B.
Ott, Wayne R.

Air/Water Surveillance and Monitoring

Barrow, David R.
Bliss, James D.
Lackey, Curtis S.
Ott, Wayne R.
Rockwell, David C.
Uchrin, Christopher

Chemical/Organic Analysis of Pollutants

Budde, Bill
Fairless, William
Heller, Stephen R.
Shew, D. Craig
Williams, Robert J.

Date Base Management

Barrow, David R.
Byram, Kenneth V.
Conger, Charles S.
Gangler, James
Hertz, Marvin
Laurie, Vernon J.
Lowrimore, Gene R.
Madsen, Mark J.
Ott, Wayne R.
Rockwell, David C.
Johnson, Richard
Jurgens, Robert B.
Knight, John E.
Lackey, Curtis S.
Scotton, John W.
Snelling, Robert N.
Worley, Donald L.

EDP Hardware

Harris, Theodore R.

Effective Use of Word/One System by OR&D

Tittle, Catherine
Webb, Ronald H.

Environmental Residuals and Land Use Modeling

Williams, Edward R.

Graphics

Borthwick, Patrick W.
Cline, David M.
Conger, Charles S.
Feldmann, Richard
Florence, Cecil E.
Goldberg, Neal
Harris, Theodore R.
Jurgens, Robert B.
Lackey, Curtis S.
Madsen, Mark J.
McCarthy, William N., Jr.
Snelling, Robert N.
Swink, Denise

Hardware Dependability

Rogers, Tommie L.

Information Storage and Retrieval Systems

Johnson, Richard
Knight, John E.
Nime, Edward J.
Scotton, John W.

Laboratory Automation

Borthwick, Patrick W.
Broadway, Jon A.
Budde, Bill
Byram, Kenneth V.
Bystroff, Roman I.
Cline, David M.
Fairless, William
Hart, John J.
Heller, Stephen R.
Laurie, Vernon J.
McCarthy, William N., Jr.
Ott, Wayne R.
Richardson, William L.
Scotton, John W.
Snelling, Robert N.
Swink, Denise
Teuschler, Jack
Williams, Robert T.

Mathematical Analysis Simulation, and Modeling

Byram, Kenneth V.
Cline, David M.
Dipert, Merlin H.
Johnson, Richard
Kinnison, Robert R.
Lowrimore, Gene R.
McCarthy, William N., Jr.
Nime, Edward J.
Ott, Wayne R.
Poole, Elijah L.
Richardson, William L.
Schuk, Walter W.
Uchrin, Christopher
Williams, Edward R.

Minicomputers

Allen, Ralph G.
Bryan, Sam D.
Bystroff, Roman I.
Cline, David M.
Dell, Robert
Feldmann, Richard
Greaves, John O.B.
Hertz, Marvin
Knight, John E.
Kyle, Kirby D.
Shew, D. Craig
Swink, Denise

Operating Systems and Analysis

Bryan, Sam D.
Cline, David M.
Dell, Robert
Greaves, John O.B.
Kyle, Kirby D.
Uchrin, Christopher

Operations Support

Rogers, Tommie L.

Programming Techniques

Byram, Kenneth V.
Gangler, James
Madsen, Mark J.

Quality Control in Laboratories

Borthwick, Patrick W.
Fairless, William
Florence, Cecil E.
Hertz, Marvin
Ott, Wayne R.
Schuk, Walter W.

Real-Time Data Acquisition and Handling

Brooks, Dorothy
Cline, David M.
Dell, Robert
Feldmann, Richard
Goldberg, Neal
Swink, Denise

Scientific/Statistical Applications and Packages

Borthwick, Patrick W.
Brooks, Dorothy
Bryan, Sam D.
Burton, Judy K.
Byram, Kenneth V.
Friedland, Michael J.
Johnson, Richard
Jurgens, Robert B.
Kinnison, Robert R.
Lowrimore, Gene R.
Male, Larry M.
Nime, Edward J.
Poole, Elijah L.
Richardson, William L.
Swink, Denise

Software Development

Harris, Theodore R.
Johnson, Maureen M.
Scotton, John W.

Spectrometry

Broadway, Jon A.
Budde, Bill
Heller, Stephen R.
McGuire, John M.
Shew, D. Craig

STORET

Allen, Ralph G.
Barrow, David R.
Bliss, James D.
Brooks, Dorothy
Byram, Kenneth V.
Florence, Cecil E.
Friedland, Michael J.
Holsomback, Will F.
Richardson, William L.
Rockwell, David C.
Tiffany, William C.

Telecommunications

Bryan, Sam D.
Conger, Charles S.
Dell, Robert
Harris, Theodore R.
Rogers, Tommie L.